KU Leuven
Biomedical Sciences Group
Faculty of Medicine
Department of Cellular and Molecular Medicine

# GENOMIC ANALYSES OF HIGH-RISK PROSTATE CANCER

Lien SPANS

Jury:

| | |
|---|---|
| Promoter: | Prof. Dr. Frank Claessens |
| Co-promoter: | Prof. Dr. Evelyne Lerut |
| Chair: | Prof. Dr. Mathieu Bollen |
| Secretary: | Prof. Dr. Jan Cools |
| Jury members: | Prof. Dr. Francesca Demichelis |
| | Prof. Dr. François Fuks |
| | Prof. Dr. Johan Swinnen |
| | Prof. Dr. Jan Cools |

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor in Biomedical Sciences

June 2014

## ACKNOWLEDGEMENTS

Dit is het dan … de laatste zinnen van mijn thesis, en in het Nederlands nog wel! Men zegt weleens dat je tijdens de 'reis' van je doctoraat een grote muur ziet opdoemen die je moet overwinnen. In mijn geval zou ik eerder spreken van meerdere muren… De voorbije vijf jaren waren voor mij als een groepsreis (ik zou het alleen niet gekund hebben!) die mij de kans gaf om mijn horizon te verbreden en ontzettend veel bij te leren. Een groepsreis, in het gezelschap van geweldige mensen, die ik hier dan ook graag zou willen bedanken.

Eerst en vooral wil ik mijn promotor Professor Frank Claessens bedanken. Frank, bedankt dat je mij vijf jaar geleden hebt opgenomen in je groeiende onderzoeksgroep. Het project is ietwat anders uitgedraaid dan we toen voor ogen hadden: 200 patiëntenstalen en de computeranalyses zou iemand anders uitvoeren. Toen puntje bij paaltje kwam en bleek dat het me met één computer niet zou lukken, stond er meteen een tweede voor me klaar! Je gaf me de tijd en het vertrouwen om me in te werken in de wondere wereld van de Linux-commando's. Je deur stond altijd voor mij open, zowel voor begeleiding, inspiratie als pep-talk. Tot slot leerde je me om niet té kritisch naar mijn eigen resultaten te kijken, om zinnen geformuleerd als half-lege glazen om te vormen tot half-volle glazen, iets wat me later zeker nog van pas zal komen.

Verder wil ik ook mijn co-promotor Professor Evelyne Lerut bedanken om mij in te wijden in de wereld van de prostaatkanker-pathologie en Gleason-scores. Bedankt dat je tijd maakte om de vele coupes en kleuringen na te kijken.

I would like to express my gratitude to the members of the Jury: Prof. Dr. Francesca Demichelis, Prof. Dr. François Fuks, Prof. Dr. Johan Swinnen and Prof. Dr. Jan Cools for your willingness to review my manuscript and for your comments and helpful suggestions.

Dit doctoraat zou er niet geweest zijn zonder de hulp en steun van de Molendo-collega's. Om de hiërarchie in stand te houden, begin ik met onze post-doc Christine: bij jou kon ik altijd terecht met mijn vele (!) vragen, met om het even welke problemen en om mijn hart te luchten. Ik mag zeker ook de QIA's niet vergeten, die ik de laatste tijd nogal vaak bij jou kwam lenen.

Rita, het kloneren van het 'kieken' van Nora is ondertussen legendarisch. Al vermoed ik dat de klonering van 'TET' je ook nog wel een tijdje zal bijblijven. Maar goed, na 6 maanden doorbijten hadden we uiteindelijk de goede constructen. Bedankt voor de toffe, ontspannende babbeltjes tussendoor, of het nu ging over wetenschap, vakanties of kleding… Hilde, in de dankwoorden van de collega's die mij zijn voorgegaan word je altijd bedankt voor het vele celcultuur-werk. Dat geldt uiteraard ook voor mij. Maar ik denk dat het vooral de Western blots zijn die je voor mij uitgevoerd hebt, waar je het meeste tijd in hebt gestoken. En telkens moest je er nog een extra filmpje tegen steken, omdat ik toch weer een ander bandje duidelijker wou zien.

Van de ex-collega's wil ik in het bijzonder Nora bedanken. Tijdens de eerste jaren van mijn doctoraat was je mijn bureaugenootje, mijn steun en toeverlaat. We babbelden nogal veel (er waren zelfs studenten die vonden dat het té veel was), en ik heb hier veel uit geleerd, al was het niet allemaal even wetenschappelijk… Stefanie, de eeuwige positivo van het labo, je slaagde er telkens weer in om een positieve kant te zien als wij het niet meer zagen zitten. Liesbeth, de sfeermaakster van het lab. Het is zo stil in't lab nu er niemand meer is die meezingt met de liedjes op de radio (en die alle teksten ervan kent).

Vanessa, het was altijd net dat tikkeltje gezelliger als we naast elkaar aan onze bench zaten en konden babbelen/roddelen tijdens het pipetteren. Je geduld en altruïsme zullen me nog lang bijblijven, net als je verwondering over weer eens een nieuw dialect-woord. Na het vertrek van Nora kreeg Michaël de moeilijke taak om bureau A te vullen. Hij heeft dit vrij letterlijk genomen en er meteen voor gezorgd dat er een mannelijk overwicht kwam door Dieter te introduceren. Michaël, bedankt voor de hulp met statistische analyses, two-way ANOVA's hebben al iets minder geheimen voor me na je uitleg. Dieter, van tijd tot tijd gaven je uitspraken mij meer inzicht in het mannelijke brein (à la vrouwen zijn ófwel slim ófwel knap). Thomas, jij brengt terug leven in onze bureau, al is het maar door het veelvuldig gebruik van de ventilator en het vele gezucht. Leen, de specialist in oestrogenen en ConA. Het lange nadenken voor je beslissingen neemt zal ik niet snel vergeten. Ik kan er nog iets van leren denk ik, al is het soms ook leuk om eens een impulsieve beslissing te nemen. Stefan, I hope you like it here in Belgium, I certainly will remember the pancakes and am curious about tasting a real Serbian specialty.

Mia en Ingrid, heel erg bedankt dat ik altijd op jullie mocht rekenen voor allerlei praktische regelingen en bestellingen.

Ik wil ook de personen bedanken die via nauwe samenwerkingen betrokken waren bij mijn onderzoek. Zeynep Kalender-Atak en Stein Aerts (Lab of Computational Biology) voor hun hulp bij de initiële exoom-analyses. Diether Lambrechts en alle mensen van zijn labo (Vesalius Research Center) voor de vele hulp met copy number analyses, Sequenom validaties en DIP-Seq experimenten. Verder ook Mark Rider en Manuel Johanns van het de Duve Instituut voor analyses van de eEF2K mutant.

Verder wil ik ook mijn vrienden bedanken voor de fijne momenten van ontspanning, onder de vorm van een etentje, een avondje uit of zelfs een vrijgezellenweekend.

Anneleen en Johan, bij jullie kon ik altijd terecht voor wat ontspanning. En sinds Thomas er is doet zijn lachende gezichtje mij alle stress vergeten.

Moeke, papa, ik ben er van overtuigd dat jullie na 5 jaar nog steeds niet goed begrijpen waar ik mij nu juist mee bezig gehouden heb. Androgenen, wat was dat ook alweer? En was het niet iets met prostaatkanker, mutaties ofzo? Toch hebben jullie mij altijd gesteund. Bedankt om mij telkens weer te motiveren/op te peppen/aan te moedigen als het even wat minder ging. Onder het motto 'moeilijk gaat ook' hebben jullie mij geleerd dat het altijd beter kan en dat ik niet te vlug mag opgeven!

Bedankt allemaal!

lien

## TABLE OF CONTENTS

## LIST OF PUBLICATIONS

### Articles in international, peer-reviewed journals

Dubois V., Laurent M.R., Sinnesael M., Cielen N., Helsen C., Clinckemalie L., **Spans L.**, Gayan-Ramirez G., Deldicque L., Hespel P., Carmeliet G., Vanderschueren D. and Claessens F. (2014). "A satellite cell-specific knockout of the androgen receptor reveals myostatin as a direct androgen target in skeletal muscle." FASEB J (epub ahead of print).

**Spans L.**, Helsen C., Clinckemalie L., Van den Broeck T., Prekovic S., Joniau S., Lerut E. and Claessens F. (2014). "Comparative genomic and transcriptomic analyses of LNCaP and C4-2B prostate cancer cell lines." PLoS One **9**(2): e90002.

Van den Broeck T., Joniau S., Clinckemalie L., Helsen C., Prekovic S., **Spans L.**, Tosco L., Van Poppel H. and Claessens F. (2014). "The role of single nucleotide polymorphisms in predicting prostate cancer risk and therapeutic decision making." Biomed Res Int **2014**: 627510.

Clinckemalie L., **Spans L.**, Dubois V., Laurent M., Helsen C., Joniau S. and Claessens F. (2013). "Androgen regulation of the TMPRSS2 gene and the effect of a SNP in an androgen response element." Mol Endocrinol **27**(12): 2028-2040.

**Spans L.**, Clinckemalie L., Helsen C., Vanderschueren D., Boonen S., Lerut E., Joniau S. and Claessens F. (2013). "The genomic landscape of prostate cancer." Int J Mol Sci **14**(6): 10822-10851.

Van Tilborgh N., **Spans L.**, Helsen C., Clinckemalie L., Dubois V., Lerut E., Boonen S., Vanderschueren D. and Claessens F. (2013). "The transcription intermediary factor 1beta coactivates the androgen receptor." J Endocrinol Invest **36**(9): 699-706.

Kerkhofs S., Dubois V., De Gendt K., Helsen C., Clinckemalie L., **Spans L.**, Schuit F., Boonen S., Vanderschueren D., Saunders P.T., Verhoeven G. and Claessens F. (2012). "A role for selective androgen response elements in the development of the epididymis and the androgen control of the 5alpha reductase II gene." FASEB J **26**(10): 4360-4372.

**Spans L.**, Kalender-Atak Z., Van Nieuwerburgh F., Deforce D., Lerut E., Aerts S. and Claessens F. (2012). "Variations in the exome of the LNCaP prostate cancer cell line." Prostate **72**(12): 1317-1327.

Helsen C., Kerkhofs S., Clinckemalie L., **Spans L.**, Laurent M., Boonen S., Vanderschueren D. and Claessens F. (2012). "Structural basis for nuclear hormone receptor DNA binding." Mol Cell Endocrinol **348**(2): 411-417.

Tanner T. M., Denayer S., Geverts B., Van Tilborgh N., Kerkhofs S., Helsen C., **Spans L.**, Dubois V., Houtsmuller A.B., Claessens F. and Haelens A. (2010). "A 629RKLKK633 motif in the hinge region controls the androgen receptor at multiple levels." Cell Mol Life Sci **67**(11): 1919-1927.

**Abstracts or posters in international meetings (first author)**

**Spans L.**, Thienpont B., Tosco L., Lambrechts D., Joniau S., Lerut E. and Claessens F. *Loss-of-function mutations in RFC1 and TET1 in primary prostate cancer.* Advances in Prostate Cancer Reseach, by the American Association of Cancer Research and the Prostate Cancer Foundation, San Diego, California, USA, 18-21/01/2014.

**Spans L.**, Kalender-Atak Z., Lerut E., Aerts S. and Claessens F. *Exome sequencing of the LNCaP and C4-2B prostate cancer cell lines.* Fifth Benelux Nuclear Receptor Meeting, Leuven, Belgium, 15/11/2012.

**Spans L.**, Kalender-Atak Z., Lerut E., Aerts S. and Claessens F. *Exome sequencing of the LNCaP and C4-2B prostate cancer cell lines.* Androgens, Helsinki, Finland, 8-10/11/2012.

**Spans L.**, Kalender-Atak Z., Lerut E., Aerts S. and Claessens F. *Exome sequencing of the LNCaP and C4-2B prostate cancer cell lines.* Oncoforum, Leuven, Belgium, 03/05/2012.

**Spans L.**, Kalender-Atak Z., Van Nieuwerburgh F., Deforce D., Haustermans K., Lerut E., Aerts S. and Claessens F. *Exome sequencing of the LNCaP prostate cancer cell line.* Keystone Symposium: Changing landscape of the cancer genome, Boston, Massachusetts, USA, 20-25/06/2011.

**Spans L.**, Kalender-Atak Z., Van Nieuwerburgh F., Deforce D., Haustermans K., Lerut E., Aerts S. and Claessens F. *Exome sequencing of the LNCaP prostate cancer cell line*. Androgens, Leuven, Belgium, 25-26/11/2010.

**Spans L.**, Kalender-Atak Z., Van Nieuwerburgh F., Deforce D., Haustermans K., Lerut E., Aerts S. and Claessens F. *Exome sequencing of the LNCaP prostate cancer cell line.* Third Benelux Nuclear Receptor Meeting, Gent, Belgium, 28/10/2010.

**Spans L.**, Van Nieuwerburgh F., Kalender-Atak Z., Aerts S., Deforce D., Swinnen J., Haustermans K., Lerut E. and Claessens F. *Exome sequencing of the LNCaP prostate cancer cell line.* Oncoforum, Leuven, Belgium, 27/05/2010.

## CURRICULUM VITAE

### Personal information


| | |
|---|---|
| Name: | Lien Spans |
| Date of birth: | May 8th 1986 |
| Place of birth: | Leuven |
| Address: | Hertstraat 37 |
| | 3090 Overijse |
| E-mail: | Lien_Spans@hotmail.com |


### Education

2009-2014: PhD in Biomedical Sciences

Laboratory of Molecular Endocrinology

Department of Cellular and Molecular Medicine, KU Leuven

*Promoter*: Prof. Dr. Frank Claessens

*Co-promoter*: Prof. Dr. Evelyne Lerut

2004-2009: Bachelor and Master in Biomedical Sciences, option Research

Title of Master thesis: SIP1 plays an essential role during the induction of hair follicles

*Promoter*: Przemyslaw Tylzanowski, Section of Reumatology, Department of Musculoskeletal Sciences, KU Leuven

1998-2004: Sint-Martinuscollege Overijse

## ABBREVIATIONS

| | |
|---|---|
| AID | Activation-induced deaminase |
| APOBEC | Apolipoprotein B mRNA-editing enzyme complex |
| AR | Androgen receptor |
| ARE | Androgen response element |
| BAM | Binary sequence alignment/map |
| BER | Base excision repair |
| Bp | Base pair |
| C | Cytosine |
| caC | Carboxylcytosine |
| CNA | Copy number alteration |
| CpG | Cytosine-guanine dinucleotide |
| CRPC | Castration-resistant prostate cancer |
| dbSNP | Single Nucleotide Polymorphism Database |
| DRE | Digital rectal examination |
| eEF2K | Eukaryotic elongation factor-2 kinase |
| EMT | Epithelial-to-mesenchymal transition |
| ERG | ETS related gene 1 |
| ETS | E26 transformation-specific sequence |
| ETV1 | ETS variant gene 1 |
| fC | Formylcytosine |
| FFPE | Formalin fixed paraffin embedded |
| FPKM | Fragments Per Kilobase per Million of mapped reads |
| Gb | Gigabases |
| GWAS | Genome-wide association study |
| hmC | Hydroxymethylcytosine |
| hmU | Hydroxymethyluracil |
| Indels | Small insertions and deletions |
| lncRNA | Long non-coding RNA |
| Mb | Megabases |
| mC | Methylcytosine |
| miRNA | MicroRNA |

| | |
|---|---|
| MLCK | Myosin light chain kinase |
| NGS | Next-generation sequencing |
| PCa | Prostate cancer |
| PCNA | Proliferating cell nuclear antigen |
| PIN | Prostatic intraepithelial neoplasia |
| PSA | Prostate-specific antigen |
| PTEN | Phosphatase and tensin homolog deleted from chromosome 10 |
| RFC1 | Replication factor C 1 |
| RP | Radical prostatectomy |
| SNP | Single Nucleotide Polymorphism |
| SNV | Single Nucleotide Variant |
| TDG | Thymine DNA glycosylase |
| TET | Ten-eleven translocation |
| TMPRSS2 | Transmembrane protease serine 2 |
| TNM | Tumor, node and metastases |
| UTR | Untranslated region |

# CHAPTER 1

# INTRODUCTION

## 1.1. PROSTATE CANCER

### *1.1.1. Epidemiology of prostate cancer*

Worldwide, prostate cancer (PCa) is the most common non-skin malignancy in men, with an estimate of 900 000 men diagnosed with PCa in 2008 (Ferlay *et al.* 2010). It is the sixth most common cause of cancer-related mortality in men worldwide, estimated to be responsible for 258 000 deaths in 2008 (Ferlay *et al.* 2010). In Europe, PCa is the most frequently diagnosed cancer and the third leading cause of death amongst men (Ferlay *et al.* 2013).

There are three well-established risk factors for PCa: increasing age, ethnic origin and genetic predisposition. Indeed, PCa is a disease of older men as the median age at diagnosis is 66 years (Howlader *et al.* 2013). Secondly, PCa is nearly 1.6 times more common in men of African descent than in white men, and black man are also twice as likely to die of PCa (Howlader *et al.* 2013). The reasons for this racial disparity are yet to be determined but may be a combination of inherited genetic, epigenetic and biochemical mechanisms. Lastly, heredity has been shown to be important since having a first-degree relative (brother or father) with PCa increases the risk for an individual by approximately two- to three-fold. If two or more first-line relatives are affected, the risk increases 5- to 11-fold (Bratt *et al.* 2002).

### *1.1.2. Diagnosis and staging of prostate cancer*

#### *1.1.2.1. Diagnosis of prostate cancer*

The main diagnostic tools include a digital rectal examination (DRE), serum concentration of prostate-specific antigen (PSA), and ultrasound-guided biopsies. PSA is a glycoprotein that is secreted by prostate epithelial cells, meaning that PSA values can be elevated both in benign and malignant conditions of the prostate. Additionally, PSA levels are not elevated in all cases of PCa. Although the PSA test was the major screening tool for PCa detection, it has drawbacks as 30-50% of patients are being overtreated, while many other patients are being undertreated (Cooperberg *et al.* 2005). Definitive diagnosis of PCa is based on histologic examination of the material obtained through biopsies.

*1.1.2.2. Staging of prostate cancer*

The TNM staging system for prostate carcinoma is based on the size of the primary tumor, the extent of invaded lymph nodes and any metastases. The primary tumor can range from organ-confined to fully invasive (T1-T4), with or without lymph node involvement (N0 or N1), and in the absence or presence of distant metastases (M0 or M1). The Gleason grade is a histological parameter which is used for grading PCa. It measures the extent of glandular differentiation and classifies tumor patterns from 1 to 5 (most to least differentiated). According to current international conventions, the Gleason score detected in a prostate biopsy consists of the Gleason grade of the dominant (most extensive) carcinoma component plus the highest grade, regardless of its extent. In radical prostatectomy specimens, both the primary and secondary Gleason grade should be reported (Epstein *et al.* 2005, Heidenreich *et al.* 2014).

*1.1.2.3. Risk stratification of prostate cancer*

D'Amico and colleagues described three risk groups that predict the risk of recurrence following localized treatment of PCa. This is based on PSA values, the Gleason score and the clinical stage (overview in Table 1.1) (D'Amico *et al.* 1998).

**Table 1.1.** Overview of the risk groups as defined by D'Amico and colleagues.

| Risk group | PSA (ng/ml) | | Gleason score | | Clinical staging |
|---|---|---|---|---|---|
| **Low risk** | ≤ 10 | And | ≤ 6 | And | T1, T2a |
| **Intermediate risk** | > 10, ≤ 20 | Or | 7 | Or | T2b |
| **High risk** | > 20 | Or | ≥ 8 | Or | T2c or higher |

Nowadays, pre-treatment risk stratification models like the Partin table, d'Amico risk groups and Kattan nomograms, which combine serum PSA level, clinical staging and biopsy Gleason score, are used in order to better predict pathological stage at radical prostatectomy and the risk of disease recurrence following definitive local treatment (D'Amico *et al.* 1998, Eifler *et al.* 2013, Kattan *et al.* 1998). Unfortunately, these stratification models have accuracies of only 75-85% and do not take into account the heterogeneity in genetic, molecular and physiological characteristics of the disease. Indeed, the reasons why some cancers progress slowly while others behave more aggressively are not well understood. At present, there are no markers to discriminate clinically relevant from indolent disease. Further improvement of the clinical management of PCa will only be possible by identifying biomarkers of PCa aggressiveness, which will not only enable a better selection of

patients who will benefit from radical treatment, but will also reduce the overtreatment of patients with indolent PCa.

### *1.1.3. Treatment of prostate cancer*

#### *1.1.3.1. Low-risk prostate cancer*

The treatments applied for localized PCa are dependent on the risk for progression that is associated with the tumor, while also taking into account the patient's age and comorbidity. Men with low-risk PCa are good candidates for active surveillance. Here, patients are followed very closely and local therapy with curative intent is initiated if there are any signs of local tumor progression (as determined by Gleason score progression at rebiopsy). This reduces the overtreatment of patients, as only 30% of men will require delayed radical intervention (Klotz *et al.* 2010).

Definitive but aggressive therapies for low-risk PCa are radical prostatectomy and radiation therapy. This active treatment is mainly recommended for patients with localized disease and a long life expectancy. Data show similar survival rates for patients with organ-confined disease who are treated with either of these methods (D'Amico *et al.* 1998). Radiation therapy can be administered externally (beam radiation therapy) or internally (brachytherapy). However, it is hard to justify such aggressive therapies with many side effects in patients with Gleason score 6, or in patients much older than 70 years (Heidenreich *et al.* 2014).

#### *1.1.3.2. Intermediate- and high-risk prostate cancer*

There is no currently accepted standard regarding the optimal treatment of men with high-risk, clinically localized PCa. Again, radical prostatectomy and radiation therapy are possibilities. Surgically removing the prostate gland in patients with intermediate-risk and high-risk PCa should always be accompanied by an extended pelvic lymphadenectomy to obtain optimal information about the extent of lymph node involvement (Joniau *et al.* 2013). This knowledge is important to counsel patients about the potential need for adjuvant treatment options (Heidenreich *et al.* 2014). When high-risk patients are primarily treated by surgery, 56% eventually require adjuvant or salvage radiation therapy or hormonal therapy (Hsu *et al.* 2007). When one chooses radiation therapy, this should always be accompanied by neoadjuvant androgen-deprivation therapy as this improves the overall survival rate (Heidenreich *et al.* 2014).

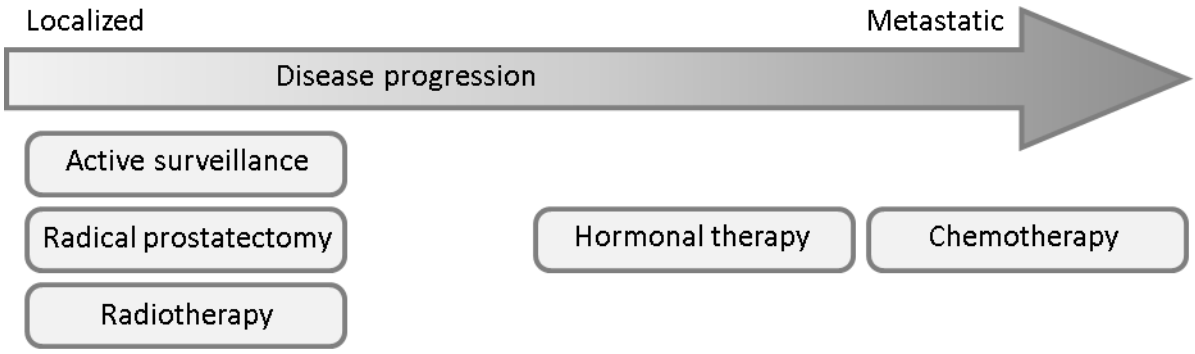



**Figure 1.1.** Overview of the most important treatments for PCa, according to the progression of the disease.

Patients diagnosed with PCa who underwent local treatment with curative intent are usually followed for at least 10 years to monitor possible recurrence indicated by a rise in serum PSA levels (Heidenreich *et al.* 2014). An overview of the most important treatments for PCa can be found in Figure 1.1.

*1.1.3.3. Advanced, relapsed and castration-resistant prostate cancer*

Hormonal therapy is a collective name for different types of drugs. Luteinizing hormone-releasing hormone (LHRH) agonists are the standard of care in hormonal therapy. Chemical castration can also be accomplished by LHRH antagonists. They result in a rapid drop of serum androgens to castration levels, which is important for the treatment of symptomatic locally advanced or metastatic disease. Finally, antiandrogens like bicalutamide can be used, although not as monotherapy (Heidenreich *et al.* 2013). Patients treated with LHRH analogues are followed by measuring serum PSA and testosterone levels. When the castration level is no longer maintained one can switch to another LHRH analogue, LHRH antagonists or add an antiandrogen (Heidenreich *et al.* 2013).

Patients treated with curative local therapies (radical prostatectomy or radiation therapy) can still relapse. When initially treated with radical prostatectomy, salvage radiation therapy is administered. When initially treated with radiation therapy, the therapeutic options are hormonal therapy or salvage radical prostatectomy (Heidenreich *et al.* 2013). Relapse after hormonal therapy is described as castration-resistant PCa (CRPC). Asymptomatic metastatic CRPC is still hormone-sensitive and can be treated with abiraterone acetate plus prednisone, while both symptomatic and asymptomatic CRPC can be treated with the chemotherapeutic drug docetaxel. The chemotherapy is mainly given to relief bone pain. Second-line treatments in CRPC can include enzalutamide or cabazitaxel plus prednisone. Because the latter therapeutic options are relatively new, there currently is no standard yet on the specific sequence of therapies (Heidenreich *et al.* 2013).

## 1.2. THE GENOMIC LANDSCAPE OF PROSTATE CANCER[1]

From a molecular point of view, cancer can result from a combination of single nucleotide variants (SNVs), small insertions and deletions (indels), rearrangements, aberrant methylation and changes in copy number which thus lead to differences in expression of oncogenes or tumor suppressor genes. A single nucleotide change is defined as a polymorphism (SNP) when two or more alleles at one locus occur in the same population, each with appreciable frequency. When the frequency of occurrence is very low, it is called a single nucleotide variant (SNV).

Recent advances in massively parallel sequencing technologies allow for the detection of all of the aforementioned changes, at a much greater sensitivity and coverage, and evenly important, also at a constantly decreasing cost. To increase the speed of analyses, one can choose for exome sequencing which allows the identification of SNVs and indels that affect the protein coding parts of the genome. Sequencing the whole genome, however, allows the additional detection of rearrangements and copy number changes. Alternatively, transcriptome sequencing not only provides data on gene expression, but can also be used to detect gene rearrangements at a lower cost than genome sequencing. In addition, these data can be linked with changes in DNA methylation obtained from high-throughput sequencing of immunoprecipitated methylated tumor DNA. Together, these techniques generate vast amounts of data that now need to be searched for correlations with disease outcome or responsiveness to specific treatments. Obtaining entire (epi)genomic and transcriptomic landscapes of PCa is contributing to the identification of targets for novel anticancer drug development. In the long run, they could assist in a better diagnosis and adequate selection of therapy tools in the settings of personalized medicine.

### *1.2.1. Copy number and gene expression changes*

Large studies on biomolecules in biological samples coming from different sources are obstructed because the tissues have undergone different preservation protocols which will affect protein, RNA and lipids (Choudhury *et al.* 2012). DNA, however, is highly stable and this makes it a preferred choice to develop reliable biomarkers. Indeed, DNA extraction enables the detection of genome-wide copy number alterations (CNAs), genome-wide SNP analyses, whole exome sequencing or even whole genome sequencing of tumor samples (Lonigro *et al.* 2011). CNAs can result in the amplification of oncogenes or the deletion of tumor suppressors, and these changes could contribute significantly to cancer etiology. Global analyses of copy number profiles of primary tumors and

[1] This chapter is based on the review 'The genomic landscape of prostate cancer' by Lien Spans, Liesbeth Clinckemalie, Christine Helsen, Dirk Vanderschueren, Steven Boonen, Evelyne Lerut, Steven Joniau, Frank Claessens, as was published in Int J Mol Sci. 2013 May 24;**14**(6):10822-51.

metastases identified recurrent aberrations associated with PCa development and progression, including broad losses of 1p, 6q, 8p, 9p and losses of large regions of chromosomes 13, 15, 18 and 22 (Reynolds 2008). Gains of 1q, 3q, 7q and 8q are also well described in PCa (Reynolds 2008). In addition, focal amplifications of the androgen receptor (AR) (Xq12) and homozygous focal deletions of PTEN (10q) and NKX3.1 (8p) are also frequent in PCa (Demichelis *et al.* 2009, Grasso *et al.* 2012, Lonigro *et al.* 2011, Taylor *et al.* 2010). A recent more comprehensive CNA study of 218 primary and metastatic tumors by Taylor *et al.* confirmed the earlier data, but added a significant role for somatic copy number increases of the NCOA2 gene, which encodes an AR coactivator (Taylor *et al.* 2010). Similarly, copy number variations of CHD1 occur in 8% of lethal CRPC samples (Grasso *et al.* 2012). CHD1 encodes an ATP-dependent chromatin-remodeling enzyme, previously reported as deregulated in PCa (Berger *et al.* 2011b). As was shown recently, some CNAs occur almost only in CRPC. Examples of these are amplification of PTK2 and YWHAZ which occur in less than 3% of primary PCas, but in 35-48% of CRPC patients (Menon *et al.* 2013).

PCa is a clinically heterogeneous disease meaning that the majority of cancer-affected prostates harbor multiple distinct primary tumor foci with different characteristics. High-resolution copy number changes from both primary tumor and different metastases revealed identical copy number changes, shared by all same-case cancer foci and defined by the same breakpoints in all multi-tumor cases (Boyd *et al.* 2012). This suggests that the genome copy number architecture was extremely homogeneous and conserved both within the primary tumor and between primary and metastatic tumors (Wu *et al.* 2012). This also indicates that metastatic PCas can have monoclonal origins and maintain the unique signature copy number pattern of the parent cancer clone (Boyd *et al.* 2012, Liu *et al.* 2009). This was confirmed recently by Haffner and colleagues, who demonstrated a clonal relationship between the primary tumor and metastases, both at the level of copy number changes and at the level of point mutations (Haffner *et al.* 2013). However, each focus will also accumulate a variable number of separate subclonally sustained genomic changes. So, although multiple tumor foci commonly arise from a single clone, this does not imply that the separate foci are biologically homogeneous. In conclusion, it is to be expected that multiple primary foci within one prostate indeed have the same genetic origin, although they may to some extent acquire distinct genetic lesions.

Another study reported an increasing percentage of the genome affected by CNAs with increasing stage, grade and diagnostic PSA levels (Cheng *et al.* 2012). This is in agreement with the study from Taylor and colleagues who reported that metastases harbor more whole chromosome, chromosome arm and focal amplifications and deletions than primary tumors (Taylor *et al.* 2010). Their study revealed three regions of recurrent copy number loss associated with the TMPRSS2-ERG fusion: two

regions spanning the tumor suppressors PTEN and TP53, respectively and a third spanning the multigenic region at 3p14 (Taylor *et al.* 2010). These data revealed distinct subgroups with substantial differences in time to biochemical (PSA) relapse. More specifically two subgroups of primary tumors were defined, those with minimal CNAs and those with substantial alterations. The latter group included most of the metastatic samples with unfavorable prognosis (Taylor *et al.* 2010). Importantly, there is no correlation between high Gleason scores and these two subgroups, indicating that histology and copy number alterations are non-overlapping features (Taylor *et al.* 2010). Hence, CNA could become useful as an additional clinical marker independent from Gleason scores.

### *1.2.2. Gene fusions*

A second type of molecular alterations occurring in cancer is the fusion or rearrangement of genes. A large number of chromosomal rearrangements were primarily discovered in leukemias, lymphomas and sarcomas (Rowley 2001). The first report on gene rearrangements in solid tumors in general and PCa in particular, however, was reported in 2005, when Tomlins and colleagues applied a statistical approach termed cancer outlier profile analysis in combination with rapid amplification of cDNA ends, thus identifying the TMPRSS2-ERG, TMPRSS2-ETV1 and TMPRSS2-ETV4 fusions in PCa samples (Tomlins *et al.* 2006, Tomlins *et al.* 2005).

#### *1.2.2.1. Detection of ETS gene fusions in PCa*

The ERG, ETV1 and ETV4 genes belong to the family of v-ets erythroblastosis virus E26 oncogenes (ETS) which encode transcription factors characterized by a highly conserved, sequence-specific DNA-binding domain, the so-called ETS domain (Donaldson *et al.* 1994). The TMPRSS2 gene encodes an androgen-regulated, type II transmembrane-bound serine protease that is highly expressed in normal prostate tissue as well as in neoplastic prostate epithelium (Vaarala *et al.* 2001a, Vaarala *et al.* 2001b). This explains why the gene fusion leads to the androgen-responsive, prostate-specific expression of these ETS transcription factors (Figure 1.2). The recurrent TMPRSS2-ETS fusion is by far the most common rearrangement described in any neoplasm, since it has been found in approximately 50% of all PCa cases examined (Perner *et al.* 2007).

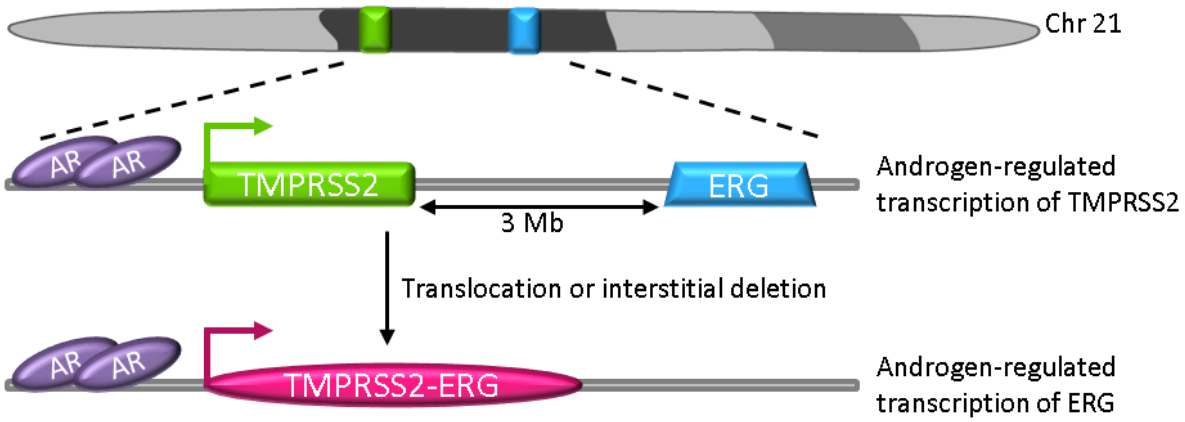



**Figure 1.2.** Mechanism of androgen-regulated overexpression of ERG. TMPRSS2 and ERG are both located on chromosome 21 in the same transcriptional orientation and only 3 Mb apart from each other. In absence of the fusion, androgen stimulation will result binding of the AR to an enhancer upstream of TMPRSS2 and hence in transcription of TMPRSS2. The TMPRSS2-ERG fusion results in the androgen-regulated overexpression of ERG.

Less common genomic rearrangements in PCa were identified later and involved SLC45A3, HERV-K, HNRPA2B1, KLK2 and C15orf21 as 5' fusion partners of ETV1 and FKBP5 as fusion partner of ERG (Pflueger *et al.* 2011, Tomlins *et al.* 2007). SLC45A3 is a prostate-specific, androgen-responsive gene that has been found fused to ERG, ETV1, ETV5 and ELK4 (Esgueva *et al.* 2010, Han *et al.* 2008, Helgeson *et al.* 2008, Maher *et al.* 2009a, Rickman *et al.* 2009, Tomlins *et al.* 2007). Recently, a SNURF-ETV1 fusion formed in conjunction with a complex rearrangement event was detected. It involves the androgen-regulated 5' fusion partner SNURF and it also led to marked overexpression of ETV1 (Weischenfeldt *et al.* 2013). An overview of all ETS gene fusions identified so far in PCa samples can be found in Table 1.2.

**Table 1.2.** Overview of ETS gene fusions detected in samples of prostate cancer patients. The list is organized according to the 3' fusion partner.

| 5' partner | 3' partner | Reference | 5' partner | 3' partner | Reference |
|---|---|---|---|---|---|
| TMPRSS2 | ERG | (Tomlins *et al.* 2005) | TMPRSS2 | ETV1 | (Tomlins *et al.* 2005) |
| HERPUD1 | ERG | (Maher *et al.* 2009b) | SLC45A3 | ETV1 | (Tomlins *et al.* 2007) |
| SLC45A3 | ERG | (Han *et al.* 2008) | C15orf21 | ETV1 | (Tomlins *et al.* 2007) |
| NDRG1 | ERG | (Pflueger *et al.* 2009) | HNRPA2B1 | ETV1 | (Tomlins *et al.* 2007) |
| FKBP5 | ERG | (Pflueger *et al.* 2011) | FLJ35294 | ETV1 | (Han *et al.* 2008) |
| TMPRSS2 | ETV4 | (Tomlins *et al.* 2006) | ACSL3 | ETV1 | (Attard *et al.* 2008) |
| DDX5 | ETV4 | (Han *et al.* 2008) | EST14 | ETV1 | (Hermans *et al.* 2008b) |
| CANT1 | ETV4 | (Hermans *et al.* 2008a) | HERVK17 | ETV1 | (Hermans *et al.* 2008b) |
| KLK2 | ETV4 | (Hermans *et al.* 2008a) | HERVK22q 11.23 | ETV1 | (Tomlins *et al.* 2007) |
| TMPRSS2 | ETV5 | (Helgeson *et al.* 2008) | FOXP1 | ETV1 | (Hermans *et al.* 2008b) |
| SLC45A3 | ETV5 | (Helgeson *et al.* 2008) | KLK2 | ETV1 | (Pflueger *et al.* 2011) |
| SLC45A3 | FLI1 | (Paulo *et al.* 2012) | FUBP1 | ETV1 | (Lapuk *et al.* 2012) |
| SLC45A3 | ELK4 | (Maher *et al.* 2009a) | SNURF | ETV1 | (Weischenfeldt *et al.* 2013) |

In general, the ETS transcription factors are considered poor therapeutic targets owing to their lack of enzymatic activity, their inaccessibility because of intranuclear activity and their dependence on interactions with other proteins to achieve specificity. Nevertheless, attempts are being made to develop compounds that interfere specifically with the function of ETS genes as transcription factors (Nhili *et al.* 2013). Alternatively, inhibitory molecules that target the TMPRSS2 promoter and/or control regions could also reduce ETS expression.

#### *1.2.2.2. Detection of non-ETS gene fusions in PCa*

Paired-end transcriptome sequencing identified genomic rearrangements involving genes of the RAF kinase pathway: SLC45A3-BRAF, AGTRAP-BRAF, ESRP1-RAF1, EPB41-BRAF and RAF1-ESRP1 (Beltran *et al.* 2012, Palanisamy *et al.* 2010). Some of the proteins encoded by these gene fusions, like BRAF, are well known drug targets, so the expression of these genes might become clinically useful in the future. Many other non-ETS gene fusions have been identified, although each of these fusions was detected only once (Lapuk *et al.* 2012, Maher *et al.* 2009a, Pflueger *et al.* 2011, Weischenfeldt *et al.* 2013, Wu *et al.* 2012). Moreover, two novel 3' fusion partners of TMPRSS2 have been identified: FKBP5 and CCDC21 (Pflueger *et al.* 2011, Weischenfeldt *et al.* 2013). Validation of the fusions involving FKBP5 led to the discovery of a complex triple fusion event with FKBP5 joined to TMPRSS2 and ERG (Pflueger *et al.* 2011). In general, the non-ETS aberrations can occur both in TMPRSS2-ERG negative and positive cancers.

#### *1.2.2.3. The role of fusion genes in the molecular pathology of PCa*

Although there are more non-ETS gene fusions identified than ETS gene fusions, most of these have been detected only once. This is in contrast to the ETS gene fusions which occur at high frequencies in PCa patients, ranging from 15 to 70% depending on the clinical cohorts investigated. Of these ETS fusions, ERG rearrangements were identified in 53% of 540 patients (Esgueva *et al.* 2010). After ERG, ETV1 is the most commonly rearranged in about 5% of the patients (Attard *et al.* 2008). Other ETS genes, such as ETV4 and ETV5 may have rearrangement frequencies at or below 1-2%.

The impact of the fusions on prognosis has been investigated in many clinical studies, but remains highly debated. A recent study involving 1039 radical prostatectomy tumors discovered that positive ERG rearrangement status is associated with younger age at diagnosis, lower serum PSA and lower prostate volume (Schaefer *et al.* 2013). In another cohort of 2800 PCas, no relation was found between ERG gene rearrangement and clinical outcome or tumor phenotype (Minner *et al.* 2011).

While the existing paradigm dictates that chromosomal rearrangements occur gradually over time, recent evidence suggests that in some cancers tens to hundreds of genomic rearrangements involving only one or a few chromosomes can occur in a cellular crisis resulting in cancer-causing lesions. This phenomenon, known as chromothripsis, was first described by Stephens and colleagues in a patient with chronic lymphocytic leukemia and several cancer cell lines (Stephens *et al.* 2011). In PCa, chromothripsis was reported one year later and was detected by the presence of triple fusion genes (Lapuk *et al.* 2012). Whole genome sequencing of seven high-risk primary tumors revealed complex inter- and intra-chromosomal events involving an exchange of 'breakpoint arms' generating a mix of chimeric chromosomes. There was, however, no loss of genetic material in contrast to what happens during chromothripsis (Berger *et al.* 2011b). These complex translocations will deregulate multiple genes in parallel and this may drive prostate tumorigenesis. To indicate the difference with chromothripsis, the latter form of rearrangements has been termed chromoplexy (Baca *et al.* 2013).

The presence of the TMPRSS2-ERG fusion in 19% of high-grade prostatic intraepithelial neoplasia (PIN) lesions adjacent to cancer foci suggests that this fusion is an early event in the development of invasive PCa (Perner *et al.* 2007). In some TMPRSS2-ERG-positive tumors, rearrangement breakpoints occur preferentially within regions containing AR and ERG DNA binding sites, while in ETS fusion-negative cells there is an inverse correlation with AR and ERG binding events, indicating alternative mechanisms for the genesis of breakpoints. This also suggests a causal link between the open chromatin structure linked to transcriptional activities at the genes involved and the mechanism of translocation (Berger *et al.* 2011b). A recent study reported on whole genome sequencing of 11 patients with early onset PCa (Weischenfeldt *et al.* 2013). Despite an overall lower number of structural rearrangements in early onset PCa compared to elderly onset PCa, they detected an increase in balanced rearrangements and a higher fraction of gene rearrangements also affecting androgen-driven genes in early onset PCa (Weischenfeldt *et al.* 2013). This contrasts with the accumulation of non-androgen-associated structural rearrangements in elderly onset PCa, most of which correspond to copy number alterations with concomitant loss of genetic material. In terms of consequences for the oncogenic process, the authors conclude that most early onset PCas involve an androgen-driven pathogenic mechanism characterized by a marked abundance of balanced structural DNA rearrangements involving androgen-regulated genes (Weischenfeldt *et al.* 2013).

The role of ERG overexpression in PCa development has been studied in transgenic mice expressing the ERG gene fusion product under androgen-regulation. These mice only develop PIN-like structures (Tomlins *et al.* 2008). The TMPRSS2-ERG fusion on its own is therefore insufficient to induce the development of invasive carcinoma, indicating that other (epi)genetic factors also contribute to the

initiation of PCa. However, the presence of the TMPRSS2-ERG gene fusion promotes PCa in both mice and humans when PTEN is concurrently lost (Carver *et al.* 2009, Casey *et al.* 2012, King *et al.* 2009).

In conclusion, although a lot is known about the gene fusions in PCa, from a clinical point of view, further classification tools, probably independent as well as dependent of the fusion status, are needed to help determine the optimal patient-tailored treatment modalities. For some cases, like BRAF-fusion positive PCa, an optimal treatment with BRAF kinase inhibitors is already available in the clinic.

### *1.2.3. Single base pair changes*

#### *1.2.3.1. Single nucleotide variants (SNVs)*

##### 1.2.3.1.1. The beginning of next-generation sequencing

Targeted resequencing of 157 genes in 80 primary tumors and metastases confirmed that the AR was the most frequently mutated gene in PCa metastases (Taylor *et al.* 2010). While it is known that alteration of the AR through mutation, gene amplification or overexpression occurs exclusively in metastatic samples after hormone therapy, alterations of the AR pathway also occur in 56% of high volume primary tumors and were confirmed in 100% of the metastases (Goh *et al.* 2012). In addition, the nuclear receptor coactivator NCOA2 had a gain of expression or mutation in 8% of primary tumors and 37% of metastases. Integration of all the mutation data with copy number alterations and transcriptome data revealed that three well-known cancer pathways were commonly altered: PI3K, RAS/RAF and RB (Taylor *et al.* 2010).

To study genomic changes in PCa, one can also study cell lines or tumors either grown *in vitro* or as xenografts in immunocompromised mice. This has the advantage that the response to cancer-directed therapeutics can be monitored, but the disadvantages are that no corresponding normal tissue or DNA is available and that a number of the genomic changes will have arisen during the culturing and hence are not relevant for the study of tumorigenesis (Kumar *et al.* 2011). Despite this, TP53 was the most frequently mutated gene in the xenografts, and pathway analysis of genes mutated in castration-resistant compared to castration-sensitive pairs of tumor lines derived from the same PCa revealed a significant enrichment of the Wnt signaling pathway (Kumar *et al.* 2011).

##### 1.2.3.1.2. Large scale genomic analyses

Berger and colleagues reported an average of 20 non-synonymous SNVs in seven high-risk primary PCas (Berger *et al.* 2011b). Only the SPTA1 gene, involved in erythroid cell shape specification, was

mutated in two out of seven tumors (Berger *et al.* 2011b). More recently, two studies explored the presence of SNVs in 112 primary tumors and 50 metastases respectively (Barbieri *et al.* 2012, Grasso *et al.* 2012). Both studies performed whole exome sequencing and reached an average of 120-fold coverage. Here, a median of 30 non-synonymous SNVs in the exome of primary PCas was detected. Most likely the higher exome sequence coverage in the latter two studies improved the detection of SNVs that are present at lower allelic fractions and thus explains the higher number of SNVs (Barbieri *et al.* 2012). Indeed, a recent genomic study with 30-40x coverage on 11 samples detected an average of only 16 non-synonymous SNVs (ranging from 3 to 55) (Weischenfeldt *et al.* 2013). Barbieri and colleagues reported twelve genes which were recurrently mutated and contained more mutations in PCa than expected by chance: PIK3CA, PTEN, TP53, SPOP, FOXA1, MED12, CDKN1B, ZNF595, THSD7B, NIPA2, C14orf49, and SCN11A (Figure 1.3) (Barbieri *et al.* 2012). Mutations in the PIK3CA, PTEN and TP53 genes were already well known to be involved in the tumorigenesis of PCa. CDKN1B was known to constrain prostate tumor growth in mice by inhibiting cell proliferation and cancer progression, but somatic substitutions had not been previously observed in this cell cycle regulator (Majumder *et al.* 2008). Several genes not previously known to undergo somatic alteration in PCa were enriched for mutations, including FOXA1, MED12, THSD7B, SCN11A, NIPA2, C14orf49 and ZNF595 (Barbieri *et al.* 2012). Some of them affect the androgen signaling axis. The transcription factor FOXA1 regulates cell proliferation and promotes tumor progression in castration-resistant PCa (Imamura *et al.* 2012). Moreover, it can act as a pioneering factor for AR binding to chromatin and the protein level in primary tumors has been associated with disease outcome (Sahu *et al.* 2011). Mutations affecting MED12 were not previously observed in PCa, but had been reported in 70% of uterine leiomyomas (Makinen *et al.* 2011). MED12 is a subunit of the mediator complex that regulates transcription by bridging DNA regulatory sequences to the RNA polymerase II initiation complex (Taatjes 2010).

The SPOP gene was mutated in 13% of the analyzed tumors (Barbieri *et al.* 2012). In additional cohorts, 6-15% of the tumors contained a SPOP mutation. In contrast to SPOP mutations detected in other cancer types, which are scattered over the entire length of the protein, the PCa mutations cluster in the substrate-binding cleft. Remarkably, the presence of a SPOP mutation was mutually exclusive with mutations in TP53, PTEN or the TMPRSS2-ERG rearrangement (Barbieri *et al.* 2012). SPOP encodes the substrate-binding subunit of an E3 ubiquitin ligase and hence is a modulator of stability for specific substrate proteins. Interestingly, it was shown recently that SPOP induces degradation of full length AR, while mutants of SPOP cannot promote AR destruction anymore (An *et al.* 2014).

Grasso *et al.* performed an exome sequencing study of metastatic biopsies of 50 lethal, heavily pre-treated CRPCs and identified nine genes that were significantly mutated (Grasso *et al.* 2012). Of these, six were already reported as recurrently mutated in PCa: TP53, AR, ZFHX3, RB1, PTEN and APC (Grasso *et al.* 2012). Three other genes were novel for PCa: MLL2, CDK12 and OR5L1. MLL2 is a histone methyltransferase that mediates H3K4 trimethylation which is recurrently mutated in multiple cancers (Natarajan *et al.* 2010). The cyclin-dependent kinase CDK12 protects cells from genomic instability through regulation of expression of DNA damage response genes (Blazek *et al.* 2011). As OR5L1 encodes an olfactory receptor, a role of this mutation in the oncogenic process is more difficult to envision. Grasso *et al.* also found that CHD1 is mutated or deleted in 8% of PCas. Using Oncomine, they detected focal deletions or mutations of CHD1 in 5.2% of 954 PCas, 96% of which were negative for the ETS-fusion. This integrated analysis identifies $CHD1^{-}/ETS^{-}$ as a novel PCa subtype (Grasso *et al.* 2012). Together, their data suggest that aberrations in AR and interacting proteins, including chromatin/histone remodelers, ETS genes and known AR coregulators including FOXA1 are common in CRPC (Grasso *et al.* 2012).

Integration of exome sequencing on primary PCa with RNA sequencing and copy number alteration revealed that the mutation rate in the mitochondrial genome was 55 times higher than that of the autosomes (Lindberg *et al.* 2012). More specifically, the electron transport chain was mutated in almost half of the tumors. Several of these mitochondrial SNVs were not yet reported, although closely related genes have been reported to be mutated in other cancers (Lindberg *et al.* 2012). For example, MLL3 was a novel gene detected to be mutated in PCa, while its close relative MLL2 has been previously reported in PCa. However, the biological and clinical relevance of these mutations needs to be documented further.

In conclusion, with over 200 PCa sequences reported, we can conclude that point mutations in PCa are not as rare as initially expected. While very few genes are recurrently affected, the mutations recur in specific signaling pathways like the androgen signaling pathway. Moreover, until studies are undertaken in substantially larger cohorts, it will be difficult to attribute significance to the different SNVs. We therefore merged the lists of SNVs detected in the aforementioned studies, both for primary tumors and metastases. The most frequently mutated genes are represented in Figure 1.3.

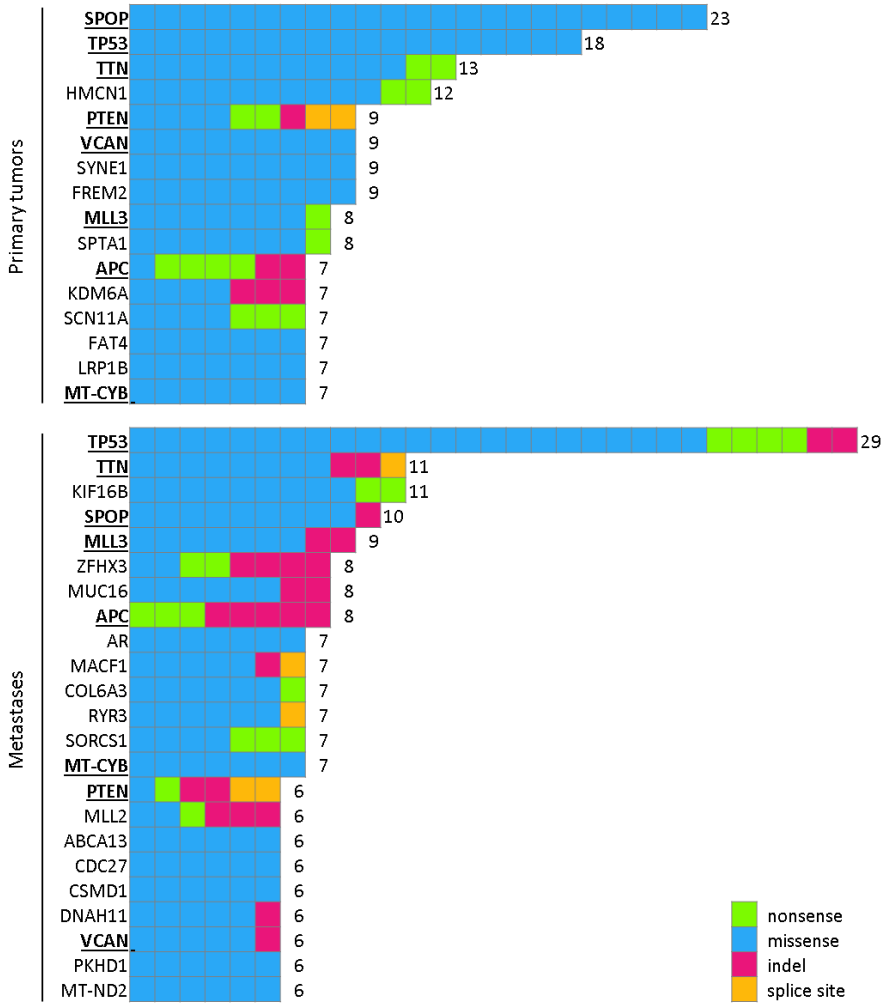



**Figure 1.3.** Overview of the most frequently mutated genes in primary and metastatic prostate cancer. Gene lists were taken from Taylor *et al.* 2010; Berger *et al.* 2011; Robbins *et al.* 2011, Barbieri *et al.* 2012; Beltran *et al.* 2012; Grasso *et al.* 2012; Lindberg *et al.* 2012; Lindberg *et al.* 2013 and Weischenfeldt *et al.* 2013. The cumulative number of mutations is given on the right. The names of genes that are recurrently mutated both in primary tumors and in metastases are bold and underlined.

#### 1.2.3.1.3. Future perspectives: Prostate cancer is a multi-focal disease

Because of the clinical heterogeneity of PCa and the presence of multiple distinct primary tumor foci, it needs to be established whether these foci are identical, similar or arose independently. Also, it has to be established whether the so-called index lesion is indeed of clinical relevance and harbors the site where metastases evolve from. In one study, four primary tumors, three of which harbored multiple foci, were sampled to verify the monoclonal or polyclonal origin of these different foci. No common SNVs were detected in the different foci of these primary tumors, indicating that the foci are independent cancers (Lindberg *et al.* 2013). Low coverage for variant calling and limited copy number analysis might explain these results that deviate from the more generally believed

hypothesis that different foci have a common origin. This was also confirmed by other studies using copy number changes.

For the development of correct DNA diagnostics, it will hence be necessary not only to study more matched primary-metastatic tumor pairs, but also to sample multiple separate tumor foci from within the same prostate. This should help to determine the molecular events that can occur during progression to advanced disease or alternatively may even help to identify less aggressive lesions.

*1.2.3.2. Single nucleotide polymorphisms (SNPs)*

The risk of developing PCa doubles for men with a first-degree relative affected by PCa and increases further with more affected relatives (Goh *et al.* 2012). This indicates that PCa is one of the most heritable cancers with up to 15% of cases linked to family history (MacInnis *et al.* 2010). Twin studies similarly suggest that up to 42% of the risk for developing PCa is linked to heritable components, indicating that the contribution of genetic factors to the development of PCa is greater than to the development of other types of common human tumors (Lichtenstein *et al.* 2000).

Genome-wide association studies (GWAS) compare the frequency of common single nucleotide polymorphisms (SNPs) throughout the entire genome (minor allele frequency > 1-5% in the population) in PCa patients and controls. In a typical GWAS, up to one million SNPs are evaluated in large cohorts of thousands of patients versus controls to determine the link between specific forms of the SNPS and the probability to develop PCa. Because only 1 or 2 million of approximately 50 million SNPs are assessed, the SNPs associated with PCa through GWAS are unlikely to be the causal genetic risk variant. However, these risk-associated SNPs are to segregate with the underlying causal variant since they are in linkage disequilibrium (Zhang *et al.* 2012). More than 70 PCa susceptibility loci explaining approximately 30% of the familial risk have been identified (Choudhury *et al.* 2012, Eeles *et al.* 2013). Additional case-control studies are generally needed to confirm the GWAS findings, as the risk of false positives is appreciable. The relative increased risk of developing the disease based on any single polymorphism discovered to date is small, generally < 1.5-fold, but risk increases with increasing number of risk alleles carried. A recent study evaluated 25 000 PCa cases and identified 23 novel PCa susceptibility loci (Eeles *et al.* 2013). Pathway enrichment of previously and newly reported susceptibility regions revealed overrepresentation in pathways regarding cell adhesion and extracellular matrix and transcriptional regulation by the AR (Eeles *et al.* 2013). Prospective GWAS studies can also evaluate rarer variants (minor allele frequency ≤ 1%) associated with PCa risk which may be more highly penetrant and carry higher relative risk.

The SNP risk markers can be located within protein-coding genes, in intergenic regions, in unannotated transcripts such as lncRNAs or miRNAs, in regulatory regions or in loci without any known genes at all. There are several potential mechanisms by which these SNPs may be associated with altered PCa risk, including genetic linkage to a coding variant in a cancer-relevant gene, changes in promoter or enhancer binding sites, changes in chromatin structure that affects expression of adjacent or distant genes, or changes in expression of noncoding RNAs (Choudhury *et al.* 2012).

The 8q24 region contains various independent PCa-susceptibility loci within a 1 Mb segment, and some of them were found to be significantly associated with other types of cancer as well, including colorectal, breast, ovarian and bladder cancer. Surprisingly, no gene has been annotated in this 1 Mb region, and its biological significance in cancer remains unclear. A possible explanation is the presence of an enhancer which physically interacts with the MYC oncogene in a tissue-specific manner (Ahmadiyeh *et al.* 2010). Similarly, the rs8072254 and rs1859961 SNPs within the 17q24 region are associated with PCa risk and modulate AR and AP-1 binding respectively, leading to an increased transcriptional activity of the prostate-specific enhancer in this block that loops to the SOX9 oncogene (Zhang *et al.* 2012).

Multiple SNPs in the promoter region of KLK3 have been associated with serum PSA levels and some have been suggested to be associated with risk of PCa (Cramer *et al.* 2003, Severi *et al.* 2006). Very recently, a SNP in the intronic region of the TERT gene at 5p15 was identified that is associated with TERT expression (Kote-Jarai *et al.* 2013). These studies demonstrate the potential interaction between genetic variants and clinical outcome.

Subjects participating in most of the GWAS studies were recruited from the general population, and thus primarily represent sporadic cancer cases. A study of SNPs in hereditary PCa indicated that at least a subset of PCa risk-related loci identified by case-control GWAS are also associated with disease risk in hereditary PCa (Jin *et al.* 2012). Several GWAS studies revealed associations of rs11672691, rs6497287 and rs1571801 with more aggressive disease, which might make them useful as prognostic markers (Amin Al Olama *et al.* 2013, Duggan *et al.* 2007, FitzGerald *et al.* 2011).

Although GWAS studies have revealed interesting aspects of PCa, the potential benefits of applying risk models based on SNPs in clinical practice are difficult. In the future, however, these genetic markers could be incorporated in clinical decision making and take part in risk models, screening paradigms and treatment recommendations.

### *1.2.4. DNA methylation*

Epigenetics is the study of heritable changes in gene expression caused by mechanisms other than those inherited via the underlying DNA sequences. Here, we focus on DNA hyper- and hypomethylation of cytosine-guanine (CpG) islands. DNA methylation can lead to gene-silencing either by inhibiting the access of target binding sites to the transcriptional activators or by promoting the binding of methyl-binding domain proteins, which interact with histone deacetylases that promote chromatin condensation into transcriptionally repressive conformations (Majumdar *et al.* 2011).

In general, overall DNA hypomethylation increases during tumor progression, while the specific hypermethylation of promoter regions of tumor suppressor genes is observed in both initiation and progression of PCa (Yegnasubramanian *et al.* 2008). The best characterized gene of which the promoter is hypermethylated in more than 90% of PCas encodes the glutathione S-transferase P1 (GSTP1) gene (Lee *et al.* 1994). More than 60 genes have been reported to be differentially hypermethylated in progressive PCa (Majumdar *et al.* 2011). Again, some of these genes have been shown to be involved in the androgen signaling pathway. Recently, the tumor suppressor miR-124 targeting the AR has been shown to be silenced by methylation in clinical PCa samples (Shi *et al.* 2012). An example of a hypomethylated gene is the plasminogen activator urokinase gene. Its increased expression is associated with higher invasive capacity of PCa cells *in vitro* and increased tumorigenesis *in vivo* (Majumdar *et al.* 2011).

Earlier epigenetic studies focused on individual or small numbers of genes. The advent of next-generation sequencing now allows profiling of methylomes, defined as the total of all DNA methylations in the whole genome. For PCa, such genome-wide studies revealed hypermethylation of homeobox or T-box genes, the FLT4, AOX1 and WFDC2 gene and dysregulation of genes involved in TNF-$\alpha$-dependent apoptosis (Kim *et al.* 2011a, Kim *et al.* 2012, Kim *et al.* 2011b, Kobayashi *et al.* 2011, Kron *et al.* 2009). Similar to the CNA conservation, the unique DNA methylation signature of the tumor/metastasis-initiating focus was shown to be closely maintained during metastatic dissemination (Aryee *et al.* 2013). The alterations in DNA methylation patterns that are associated with phenotypic changes in gene expression have a surprising strong tendency to be maintained within metastases in an individual patient (Aryee *et al.* 2013).

Methylome analysis of 51 primary PCas identified 147 000 cancer-associated epigenetic alterations, of which 58% were hyper- and 42% hypomethylated (Borno *et al.* 2012). Tumors without the TMPRSS2-ERG fusion contain more differentially methylated regions than fusion-positive tumors,

suggesting a more pronounced role for epigenetic mechanisms in fusion-negative tumors (Borno *et al.* 2012).

Clearly, in combination with DNA mutations and gene fusions, DNA methylation markers hold great promise as clinically useful diagnostic or prognostic parameter. One major open question is whether specific subtypes of the disease might be identified by combinations of hyper- and hypomethylation events. This is one of the topics that might be answered by more comprehensive, genome-wide studies.

### *1.2.5. Non-coding RNAs*

#### *1.2.5.1. MicroRNAs*

MicroRNAs are small, non-coding RNA molecules that bind to the 3' untranslated region (UTR) of mRNA. This binding effectively silences translation by blocking access to the ribosome or by marking the target mRNA for degradation. Genes encoding miRNAs are found as independent entities or within introns of other genes, within repetitive genomic elements or within transposable element sequences.

At present, more than 100 miRNAs have been reported to be deregulated in PCa. There are however many conflicting results in the literature which is likely due to the still immature technology to capture and quantitate miRNAs and the contamination of normal cells in the tumor samples (Carlsson *et al.* 2011, Fuse *et al.* 2012, Sun *et al.* 2009, Volinia *et al.* 2006). Despite these inconsistencies, all studies confirm the widespread dysregulation of miRNAs in PCa. Moreover, a subset of these have been experimentally shown to be involved in the initiation, progression from androgen-dependent to androgen-independent stage, invasion and/or metastasis of PCa (reviewed in (O'Kelly *et al.* 2012)). A growing number of miRNAs is being identified as interfering with the AR pathway. A gain-of-function screen in PCa cell lines identified 13 unique miRNAs that influence the level of AR in these cells (Ostling *et al.* 2011). On the other hand, androgens control the up-regulation of miR-125, miR-21 and miR-141 and consequently the down-regulation of their respective target mRNAs (Ribas *et al.* 2009, Shi *et al.* 2007, Waltering *et al.* 2011).

Oncomirs are miRNAs that are dysregulated in cancer. Examples of miRNAs that show decreased expression in PCa compared to normal prostate tissue are miRNA-143, -145 and -200. The mRNA targets for these miRNAs are being discovered, and it seems that the miRNA down-regulation results in epithelial-to-mesenchymal transition and enhanced tumor cell invasion and migration (Gregory *et al.* 2008, Korpal *et al.* 2008, Peng *et al.* 2011). Conversely, miRNAs that show increased expression are miRNA-21 and miRNA-125b as they play important roles in resistance to apoptosis (Lee *et al.*

2005, Li *et al.* 2009c). Here too, although some downstream mRNA targets are being discovered, most of them remain poorly understood. So far, many studies examining the role of miRNAs are associative and rely on PCa cell lines as surrogates for clinical response. It is thus envisaged that the focus will shift towards clinically relevant studies both in animals and humans to provide a better understanding of the working mechanisms of miRNAs (O'Kelly *et al.* 2012). Alternatively, studies in humans could pinpoint those microRNAs that can be used as a prognostic marker. In this way, miR-221 was discovered to be progressively down-regulated in primary PCa and metastasis (Spahn *et al.* 2010). This down-regulation is associated with Gleason score, tumor progression and clinical recurrence during follow-up (Spahn *et al.* 2010).

Interestingly, certain microRNAs are not only elevated in the prostate tumor, but also in the circulation (for example in the exosomes) of the patient, suggesting they act similar to hormones and might play a role in priming the site of metastasis. An example is the up-regulation of miR-375 which can predict biochemical relapse, with high expression being associated with an unfavorable outcome (Selth *et al.* 2012). Circulating miRNAs open up the possibility of their development into diagnostic tools. Ultimately, miRNAs could be used to predict outcome and response to treatment or even be targets of treatment themselves.

*1.2.5.2. Long non-coding RNAs (lncRNAs)*

LncRNAs share common traits with mRNA because they are mostly transcribed by RNA polymerase II, they are capped, polyadenylated and spliced but they do not contain an open reading frame. Most of the lncRNAs identified to date display overexpression in PCa samples. A few lncRNAs seem to be prostate-specific: PCA3/DD3, PCGEM1, PRNCR1 and PlncRNA-1.

The best documented is PCA3 which was discovered as a differentially expressed RNA (DD3). It is expressed exclusively in the prostate, is highly overexpressed in PCa and detectable in urine samples of PCa patients (Bussemakers *et al.* 1999). PCA3 has been conclusively shown to be a better biomarker for PCa in biopsy samples than PSA. PCA3 and PSA together are an even better predictor of PCa (Wang *et al.* 2009). This led to the FDA approval of a PCA3 urine test as diagnostic aid to decide on repeated biopsy testing.

Other lncRNAs are still in the early discovery phase. PCGEM1 also encodes an androgen-regulated lncRNA that is expressed exclusively in glandular epithelial cells of both normal and tumor specimens of human prostate (Srikantan *et al.* 2000). In patients, tumor-associated overexpression of PCGEM1 was detected in 84% of the samples (Srikantan *et al.* 2000). Probably the least characterized lncRNA is PRNCR1 (prostate cancer non-coding RNA 1), an approximately 13 kb intron-less non-coding RNA

transcribed from the 8q24 region (Chung *et al.* 2011). In a small cohort, the PRNCR1 expression was up-regulated in half of the PCa samples as well as in the precursor PIN lesion. Finally, PlncRNA-1 was found to be overexpressed in 11 out of 16 PCa samples and a knockdown resulted in decreased cell viability, increased apoptosis and a decrease of AR mRNA and protein (Cui *et al.* 2013).

Transcriptome analyses of a cohort of 81 prostate tissues led to the discovery of 121 unannotated PCa-associated lncRNA transcripts (PCATs). Similar to the gene signatures, changes in the levels of these transcripts are being studied for their use as diagnostic tool. In a first study, their expression levels accurately discriminate benign, localized tumor and metastatic prostate samples (Prensner *et al.* 2011). One of these transcripts, PCAT-1, seems to be a prostate-specific transcriptional repressor that regulates cell proliferation and that may hence have an important role in PCa progression (Prensner *et al.* 2011). Similar approaches might contribute to identify additional disease-associated lncRNAs that may further improve the stratification of cancer subtypes.

### *1.2.6. A role of AR in PCa*

It stands without debate that the AR protein is a crucial transcription factor in normal as well as diseased prostate, and that it plays a pivotal role both during development and progression of PCa. As a transcription factor, it controls proliferation as well as differentiation of prostate cells by regulating processes at multiple levels (proteins, miRNAs and lncRNAs) (Balk *et al.* 2008). The AR gene itself is a target for (de-)methylation, amplification and mutagenesis that lead to gain of function. Many of the events described above not only affect the AR gene itself, but the entire pathway, for example by disturbing AR cofactors. Since its dramatic effects on chromatin structure, the AR is now also a prime suspect to play a role in many of the genomic events like translocations. It is even becoming more and more clear that part of the genetic predisposition to PCa also involves the AR at some level, for example through the effect of SNPs on androgen-response (Clinckemalie *et al.* 2013). All these processes are illustrated in Figure 1.4.

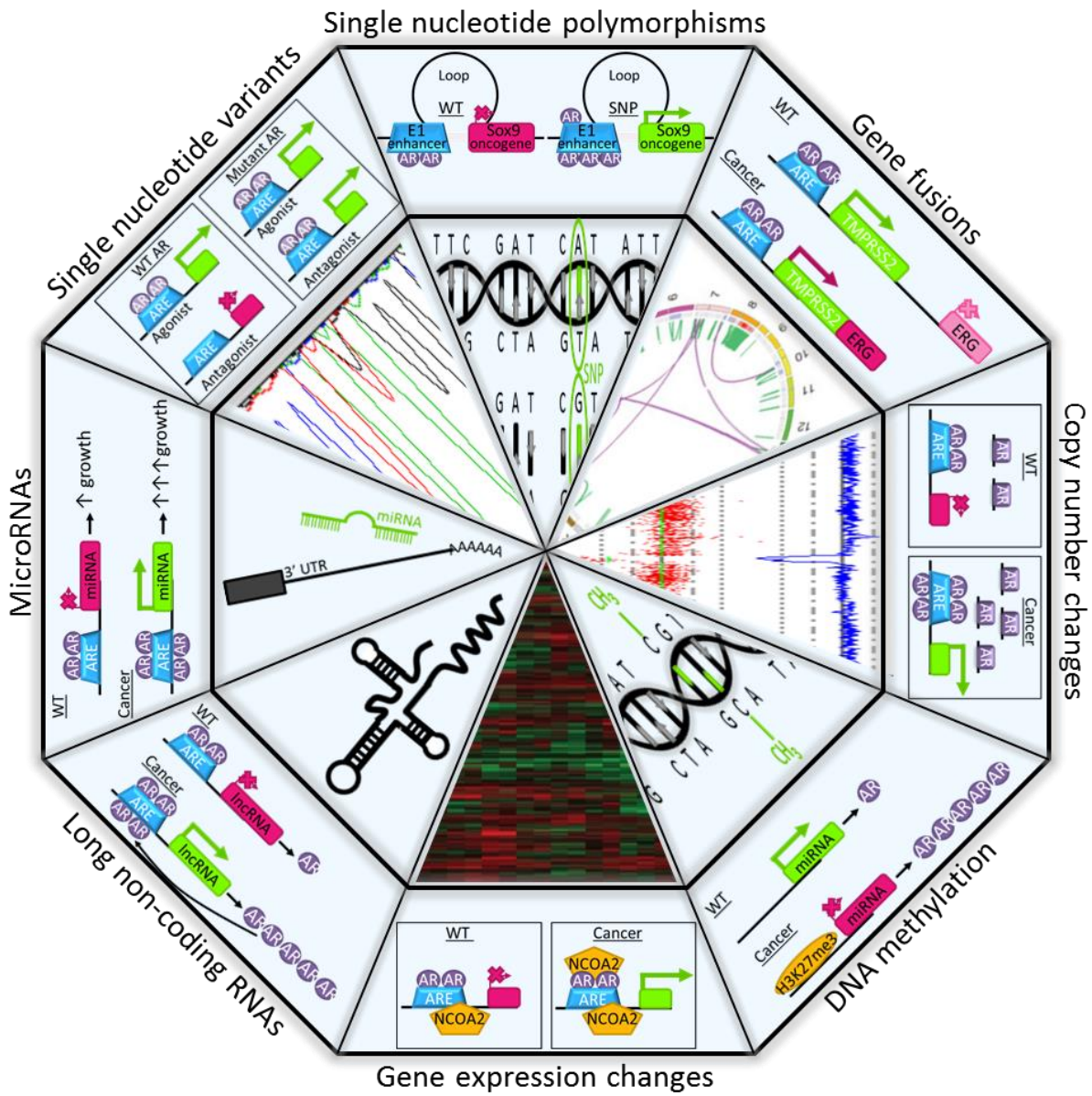



**Figure 1.4.** The genomic landscape of PCa. The integrated analysis of DNA, RNA and methylation data obtained with next-generation sequencing will help elucidate all relevant (epi)genetic changes in cancers (inner). Involvement of the androgen receptor in PCa tumorigenesis and progression can be demonstrated at all levels (outer).

## 1.3. CELL LINES AS MODEL FOR PROSTATE CANCER

### *1.3.1. Use of different prostate cancer cell lines*

Recently, a debate is ongoing about the relevance of cell lines: are they or are they not representative of the original tumor? To what extent do the cells actually resemble their tissue of origin after years in culture? Nevertheless, cell lines still play an essential role in biomedical research as they are cheap and ensure sufficient material for long-term use. Despite numerous attempts to obtain PCa cell lines, only a handful of human PCa cell lines have been generated, of which the most commonly used were isolated from metastatic lesions rather than primary tumors (Table 1.3). Moreover, each model system displays its own characteristics.

**Table 1.3.** Overview of the PCa cell lines that will be used in this doctoral thesis.

| Name | Origin | AR protein | PSA protein | References |
|---|---|---|---|---|
| C4-2B | Bone metastasis of LNCaP in nude mice | Mutated | Positive | (Thalmann *et al.* 1994) |
| CWR22Rv1 | Xenograft from primary PCa | Mutated | Positive | (Sramkoski *et al.* 1999) |
| DU145 | Brain metastasis | Negative | Negative | (Stone *et al.* 1978) |
| LNCaP | Lymph node metastasis | Mutated | Positive | (Horoszewicz *et al.* 1980) |
| PC-3 | Vertebra metastasis | Negative | Negative | (Kaighn *et al.* 1979) |
| VCaP | Vertebra metastasis | Wild type | Positive | (Korenchuk *et al.* 2001) |

The first human PCa epithelial cell lines to be established were LNCaP, PC-3 and DU145. Even now, these three cell lines and their derivative sublines are still used in the majority of studies. As PC-3 and DU145 cells do not express significant levels of AR and PSA, they cannot be used to study androgen signaling or castrate-sensitivity. Although LNCaPs are androgen-responsive and produce PSA, they express a mutated AR (T877A), which results in broader ligand-specificity (Veldscholte *et al.* 1990). LNCaP and C4-2B cells represent a progression model that mimics advancing disease from poorly tumorigenic, non-metastatic in LNCaP, to metastatic and castration-resistant in C4-2B.

A brain metastasis with moderately differentiated adenocarcinoma gave rise to the DU145 cells. The PC-3 cells were established by using poorly differentiated metastatic tumor tissue from a lumbar vertebra. Both patients had previously been castrated as treatment for the PCa.

More recently, novel cell lines have been established, among which the VCaP cell line. These cells were derived from a vertebral bone metastasis from a man with hormone refractory PCa (Korenchuk *et al.* 2001). The tissue was first heterotransplanted into immune-deficient mice and later harvested for *in vitro* culture. These androgen-sensitive cells express higher levels of wild type AR than LNCaP

cells. Moreover, they harbor the TMPRSS2-ERG fusion, which makes them an appropriate system to investigate the functional significance of this rearrangement (Perner *et al.* 2006).

CWR22 cells have been propagated from an epithelial xenograft derived from a primary prostate carcinoma from a patient with osseous metastases. CWR22R is a relapsed xenograft from castrated mice carrying the androgen-dependent CWR22. Also, the CWR22Rv1 cell line is the first variant of CWR22R cells that could be propagated in cell culture (Sramkoski *et al.* 1999).

### *1.3.2. Development of LNCaP cells*

LNCaP cells were grown out of small tissue fragments obtained by needle aspiration biopsy of a metastatic lesion in the left supraclavicular lymph node of a 50-year-old Caucasian male (Horoszewicz *et al.* 1980). This patient suffered from a rapidly progressing PCa with minimal and temporary response to hormonal therapy and no response to chemotherapy.

The patient was diagnosed with a moderately differentiated PCa based on needle biopsy of the prostate and retroperitoneal lymph node exploration. Treatment with oral estrogens could not prevent the development of bone metastases six months after diagnosis. Castration only resulted in a mild and temporary response since the patient presented with a hard enlarged prostate and pain in his right flank five months after castration. One month after the start of chemotherapy, a needle biopsy was taken from a palpable left supraclavicular lymph node which was the source of LNCaP cells. Finally, 18 months after diagnosis, the patient died.

The LNCaP cells express many of the prostate-specific genes and are AR positive. For this reason, it has been used in many preclinical studies. Several sublines of these LNCaP cells, among which the C4-2B cells, were established to represent CRPC.

### *1.3.3. Development of C4-2B cells*

The C4 subline was derived through the coinoculation of two cell lines: one million LNCaP cells were injected subcutaneously into a male athymic nude mouse, together with one million cells of the human osteosarcoma cell line MS. The host was castrated 8 weeks after injection and a single tumor specimen was excised 4 weeks later. This specimen was used as the source for the generation of a second generation cell line, C4-2, maintained for 12 weeks in a castrated host after subcutaneous injection together with the MS fibroblasts. Finally, C4-2 cells were inoculated orthotopically in a castrated mouse. At the time when metastases were detected, the mouse was sacrificed and tumor

tissue was harvested from the bone for subsequent derivation of the C4-2B subline (Thalmann *et al.* 1994). A schematic overview of the entire process can be found in Figure 1.5.

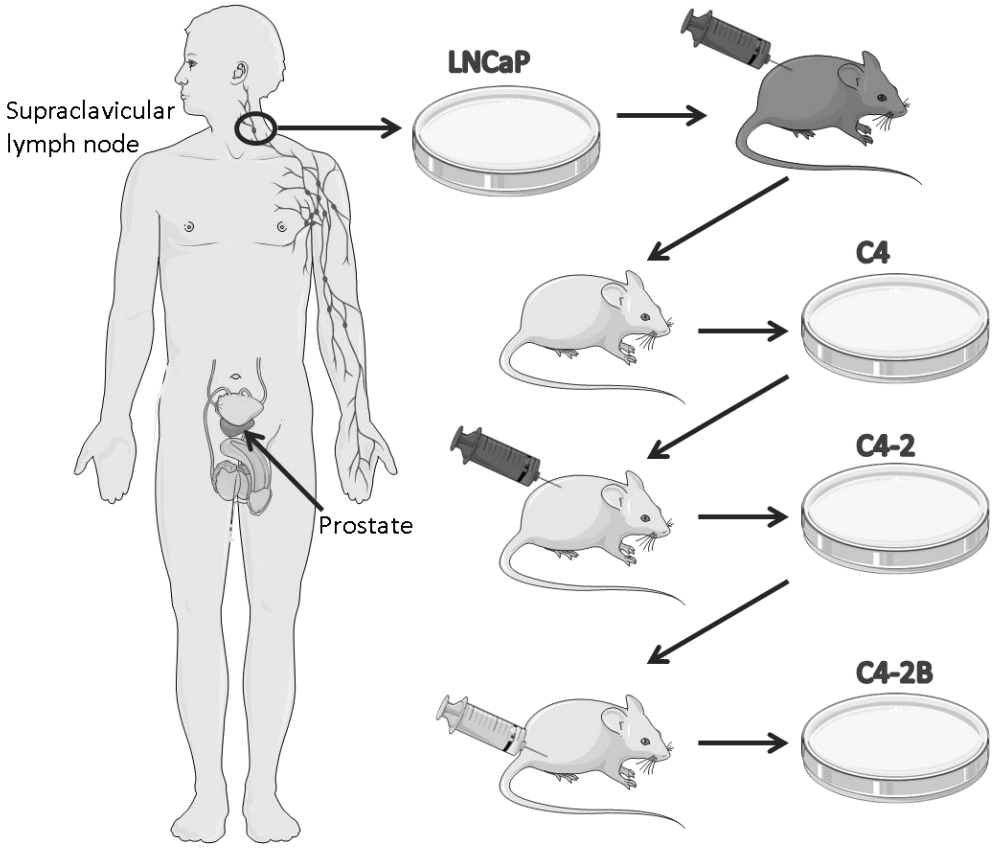



**Figure 1.5.** The development of LNCaP and C4-2B PCa cell lines. LNCaP cells were isolated from a supraclavicular lymph node metastasis. The C4-2B cell line is derived from a LNCaP tumor grown in castrated mice. Intact and castrated mice are represented in dark and light gray respectively. A dark syringe represents subcutaneous injection of PCa cells with human fibroblast. A light syringe indicates orthotopic injection of PCa cells only.

## CHAPTER 2:

## AIMS OF THIS STUDY

Cancer cell lines are commonly used as laboratory resources to study basic molecular and cellular biology. For PCa, LNCaPs are the most commonly used cells, as they have been cited in over 6200 papers (number of Pubmed hits with the term 'LNCaP' in March 2014). However, the interpretation of experiments on these cells is hampered by the absence of a known mutational status of the genes involved. **The first aim** of this project was therefore to provide an extensive database of genetic variations in the exome of LNCaP cells. We focused on missense and nonsense single nucleotide variants (SNVs) and short insertions and deletions (indels). We also tried to answer the question whether different LNCaP cell lines which are known to be quite variable between laboratories, are genetically unstable, heterogenous and/or heterozygous.

The LNCaP/C4-2B progression model mimics the phenotypic and genotypic changes often observed in PCa patients when evolving from androgen-sensitive to metastatic, castration-resistant PCa. **The second aim** was to characterize both LNCaP and C4-2B cell lines more thoroughly. We therefore performed exome and transcriptome sequencing, from which we could derive SNVs and differential expression patterns. Based on these results, we tried to detect some molecular differences that could explain the increased metastatic capacity of the C4-2B cells. In addition, these analyses were used as training for aim 3.

PCa is a very heterogeneous disease, ranging from indolent to very aggressive tumors. It still remains a clinical challenge to differentiate both forms of PCa. A better molecular profiling of the tumors should enable a better classification of the disease, should unveil the signaling pathways that are involved in the carcinogenesis and ultimately could provide information that could direct a more personalized treatment. One approach in doing so is to study the contribution of somatic base pair substitutions to the oncogenic process. At the onset of this study little was known about PCa mutations. In the meantime, several reports have been published. However, new substitutions are discovered with almost every new case that is analyzed. **The third aim** of this project was therefore to perform whole exome sequencing of tumors from high-risk PCa patients and identify the tumor-specific SNVs and hence the genes that are disrupted. The tissues were collected within the frame of the PEARL (ProstatE cAncer Research team Leuven) consortium. Finally, the functional consequences of some of the SNVs have been studied, together with their potential influence on the initiation and/or aggressiveness of PCa. This would then result in a deeper insight in the developmental mechanisms of PCa.

## CHAPTER 3:

## VARIATIONS IN THE EXOME OF THE LNCAP PROSTATE CANCER CELL LINE

Lien Spans, Zeynep Kalender-Atak, Filip Van Nieuwerburgh, Dieter Deforce, Evelyne Lerut, Stein Aerts and Frank Claessens

Prostate; **72**(12); 1317-27; 2012

The work presented in this chapter has been conducted by the PhD candidate.

### 3.1. ABSTRACT

**Background.** The LNCaP cell line is widely used as a model for prostate cancer. However, information on protein-changing mutations, genetic heterogeneity and genetic (in)stability is largely lacking for these cells.

**Methods.** Next-generation sequencing of the LNCaP exome revealed many single nucleotide variants (SNVs). To help identify the mutations that are most likely drivers of the oncogenic process, we developed an *in silico* protocol, which can be adapted for other exome analyses.

**Results.** We detected 1802 non-synonymous SNVs and 218 small insertions and deletions in the LNCaP exome. We confirm the known mutations in the androgen receptor and the PTEN gene, but most other mutations remained undescribed until now. The presence of 38 out of 42 SNVs was confirmed in monoclonal as well as in polyclonal LNCaP derivatives. Moreover, most variants were also detectable in LNCaP mRNA.

**Conclusions.** We provide an extensive database of genetic variations in the protein-coding part of the genome of LNCaP cells, which should be taken into consideration when using LNCaP cells or its derivatives as models for prostate cancer. From the analysis of several LNCaP derived cultures and clones, we can confirm that the cell line is heterozygous for a large number of variants and that both the variant and the wild type allele can be simultaneously expressed as mRNA. The fact that the SNVs in the E-cadherin, CDK4, Notch1 and PlexinB1 genes are absent in some of the subclones strongly indicates a degree of genetic instability.

## 3.2. INTRODUCTION

The LNCaP PCa cell line was derived from a biopsy taken from a lymph node of a fifty-year-old Caucasian male with metastasized PCa (Horoszewicz *et al.* 1980). These androgen-sensitive cells have been used in work leading to at least 5000 publications over the last three decades and are the most frequently used *in vitro* model for basic as well as preclinical studies of PCa. A wide range of biological information such as its transcriptome, karyotype and limited analyses of mutations has been reported (Bainbridge *et al.* 2006, Pan *et al.* 1999, van Bokhoven *et al.* 2003). However, to our knowledge, there are no data available on the exome or the complete exon content of these LNCaP cells.

Here, we re-sequenced the complete exome of the LNCaP PCa cells with the aid of next-generation sequencing, in search for mutations that lead to amino acid substitutions or to small exonic insertions and deletions. We have tested the quality of our data by re-sequencing genomic PCR products encompassing over 40 exons. The mutations will be discussed in light of the recently reported genomic data on PCa (Berger *et al.* 2011b, Kumar *et al.* 2011, Robbins *et al.* 2011, Taylor *et al.* 2010). To determine which of the mutations would be the best candidate driver mutations, we developed an *in silico* prediction algorithm which could be useful for the analyses of (prostate) cancer genomic data.

The biological characteristics of the LNCaP cells, like for many other cell lines, are known to vary between different laboratories or even between different time points during culture (Esquenet *et al.* 1997). It is unknown whether these differences are due to heterogeneities within the cell line, phenotypic changes like adaptations to different culture settings, or to genotypic drift by relatively high mutation rates combined with culture-induced selection. We will discuss these possibilities by verifying the presence of a subset of the SNVs and SNPs in LNCaP cultures from different origins.

## 3.3. MATERIALS AND METHODS

### *3.3.1. Cell culture*

Cell lines (LNCaP, PC-3, DU145, VCaP and CWR22Rv1) were obtained from the American Type Culture Collection. LNCaP (UK) and LNCaP TR2 were kindly provided by Dr. C.L. Bevan (Imperial College, London, UK). LNCaP (NL) and C4-2B were obtained from Dr. J. Trapman (Josephine Nefkens Institute, Erasmus University Medical Centre, Rotterdam, The Netherlands) and Dr. M.R. Stallcup (University of Southern California, LA, CA, USA) respectively. The passage number of the LNCaP cells, LNCaP (UK)

and C4-2B cells is 48, 12 and 42 respectively. LNCaP, C4-2B and CWR22Rv1 PCa cells were grown in Roswell Park Memorial Institute medium (RPMI, Gibco, Invitrogen, Gent, Belgium) containing 2 g/l glucose supplemented with 10% heat-inactivated fetal calf serum (FCS). VCaP cells were grown in Dulbecco's Modified Eagle's Medium (DMEM, Gibco) with 4.5 g/l glucose. DU145 cells were grown in DMEM containing 10% FCS and 4 µg/ml insulin. PC-3 cells were grown in Ham's F12 (Gibco)/DMEM (1:1) supplemented with 10% FCS. All cells were grown at 37°C and 5% $CO_2$ and media were supplemented with 100 µg/ml streptomycin and 100 µg/ml penicillin.

### *3.3.2. DNA isolation and whole-exome capture sequencing*

High-molecular weight DNA was extracted from cultured LNCaP cells using the GenElute Mammalian Genomic DNA Miniprep Kit (Sigma-Aldrich, Bornem, Belgium). The DNA was further purified using ethanol precipitation with ammonium acetate. The concentration was quantified using a Nanodrop (Thermo Fisher Scientific, Erembodegem-Aalst, Belgium), Quant-iT PicoGreen (Invitrogen) and Bioanalyzer (Agilent Technologies, Diegem, Belgium). Five microgram of DNA was fragmented using an ultrasonic solubilizer (Covaris, KBiosciences, Utrecht, The Netherlands) (10% duty, intensity of 5 with 200 cycles per burst for 180s) for the construction of a 200 base pair fragment DNA library. Whole-exome capture was performed using the SureSelect Human All Exon System (Agilent Technologies) according to the manufacturer's instructions. We generated 100 base pair long paired-end reads using the GAIIx sequencer (Illumina, Eindhoven, The Netherlands). We used one lane of a flow cell for the sample. The produced image data were converted to intensity files and were processed through the Firecrest and Bustard algorithms provided by Illumina to call the individual sequence reads.

### *3.3.3. Bioinformatics: Sequence alignment, removal of PCR duplicates and detection of SNVs and indels*

Quality control was performed using FastQC software (version 0.4) (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/). Paired-end reads were aligned to the human reference genome (hg19, NCBI Build 37) using Bowtie (version 0.12.4) and Burrows-Wheeler Aligner (BWA) (version 0.5.6) (Langmead *et al.* 2009, Li *et al.* 2009a). Processing of the aligned reads was done using SAMtools (version 0.1.7) (Li *et al.* 2009b). Because there were duplicated reads which were generated during the PCR amplification process, paired-end reads that aligned to the same genomic positions were removed using Picard (version 1.22) (http://picard.sourceforge.net/).

To detect SNVs, we used the VarScan program (version 2.1) with default parameters (Koboldt *et al.* 2009). Variant calls were filtered according to different criteria using the Galaxy website (see Results) (Blankenberg *et al.* 2010, Goecks *et al.* 2010). For the detection of indels, we used the Dindel software (version 1.01) (Albers *et al.* 2011). Default options were used, together with the criterion that each candidate indel had to be seen in at least two supporting reads. Reads were visualized with Integrative Genomics Viewer (IGV, version 1.4.2) (Robinson *et al.* 2011), while SNVs and indels are represented in a Circos plot (version 0.52) (Krzywinski *et al.* 2009). BEDtools (version 2.6.1) was used to calculate coverage information (Quinlan *et al.* 2010).

### *3.3.4. In silico prediction of impact of SNVs on protein function*

Ranking A is a combination of the predictions of three independent software programs (SIFT, PolyPhen2, and MutationAssessor) (Adzhubei *et al.* 2010, Kumar *et al.* 2009, Reva *et al.* 2011). Each of these programs uses different calculations which result in an internal score that reflects the probability that an amino acid change will be accepted at a given position of a protein sequence. We therefore first normalized the internal score of MutationAssessor to values between 0 and 1, and took the complement of the SIFT probability as the normalized score of this tool. The average of the three normalized scores resulted in Ranking A. Ranking B was obtained using the Endeavour software and includes functional and pathway annotations, protein-protein interactions, expression levels and literature (Tranchevent *et al.* 2008). We used two different training sets for Endeavour. A first one is the KEGG pathway for PCa. A second one is a list of 115 genes that we found in literature to be involved in PCa (development). Again, the result of both training sets was combined to obtain one Ranking B. Both Ranking A and B assign a rank to each gene. By averaging both ranks, we obtained the final ranking.

### *3.3.5. Sanger sequencing to validate non-synonymous substitutions*

Primers for validation were designed by targeting regions immediately flanking the predicted variant using the NCBI Primer-Blast (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) and obtained from Integrated DNA Technologies (Haasrode, Belgium). Polymerase chain reactions were performed following standard protocols using Taq DNA polymerase (Fermentas GmbH, Thermo Fisher Scientific). Products were run on an agarose gel and assessed for purity and size. The remaining products were purified using the GeneJet PCR Purification kit (Fermentas GmbH). Sanger sequencing was performed at LGC Genomics. Sequence trace files were analyzed using Chromas Lite.

## 3.4. RESULTS

We performed an exome re-sequencing study of the LNCaP PCa cell line. Exonic fragments were enriched from sonicated genomic DNA using Agilent's SureSelect technology for targeted exon capture, which targets 37 Mb of sequences from exons and their flanking regions in ~20000 genes. Sequencing with the Illumina GAIIx platform resulted in 25348495 raw 100 base pair long, paired-end reads (5.07 Gb of sequence). These reads were then aligned to the human reference genome (build 37) using BWA and Bowtie (see Methods) (Langmead *et al.* 2009, Li *et al.* 2009a).

### *3.4.1. Detection and filtering of SNVs*

On average, 93% of all exonic nucleotides in the SureSelect kit (target regions) were covered with at least one sequencing read, while 78% of these nucleotides were covered at least 12 times, the latter of which was used as a cutoff in our further analyses. The average sequencing read depth is 43x in target regions. Using the Bowtie or BWA aligner followed by the VarScan software, 193595 and 306844 SNVs were predicted, respectively (Koboldt *et al.* 2009, Langmead *et al.* 2009, Li *et al.* 2009a). All the variant calls were filtered using a combination of criteria that reduced the likelihood that a sequencing error was identified as a single nucleotide variant (Table 3.1). These filtering criteria required that at least 12 reads covered a potential variant, that all reads had a Phred-quality score of at least 30 and that the frequency of the non-reference allele was at least 30%. Variants passing these filters were annotated by Ensembl using the Perl API. We further selected those SNVs that fall into protein-coding regions and would result in a change of the amino acid sequence. Mapping to dbSNP (version 131) allowed us to separate known SNPs from *de novo* variants. We could not determine which of the remaining SNVs were somatic or tumor-specific, since no matched normal tissue from the corresponding patient is available. Next, the remaining SNVs were investigated to remove all strand biases as these are potential sequencing errors. Finally, only SNVs that were found after alignment with Bowtie as well as with BWA were taken into consideration. Using this approach, 1685 non-synonymous coding SNVs and 117 SNVs that introduce a stop codon in protein-coding regions were found in 1610 different genes (Table 3.1). Of these, only 98 SNVs or 5.4% are homozygous (with a mutation allele frequency of at least 70%). We observed a ratio of non-synonymous to synonymous changes of 2.1:1, which is not higher than the ratio of 2:1 predicted for non-selected passenger mutations (Bardelli *et al.* 2003).

**Table 3.1.** Filters used to identify point mutations in the LNCaP exome

| *Number of SNVs* | *BWA* | *Bowtie* |
|---|---|---|
| Predicted | 306 844 | 193 595 |
| With ≥ 12x coverage | 286 603 | 177 179 |
| With ≥ 30 Phred-quality | 203 710 | 130 941 |
| With ≥ 30% mutation allele frequency | 48 788 | 42 346 |
| In coding regions | 17 567 | 17 211 |
| Non-synonymous coding or gained stop codon | 8 900 | 8 671 |
| Not present in dbSNP131 | 1 997 | 1 987 |
| Without strand bias | 1 957 | 1 959 |
| In common | 1 802 | |

Mutation allele frequency denotes the number of reads containing the mutation divided by the total number of reads. The list of SNVs is available upon request.

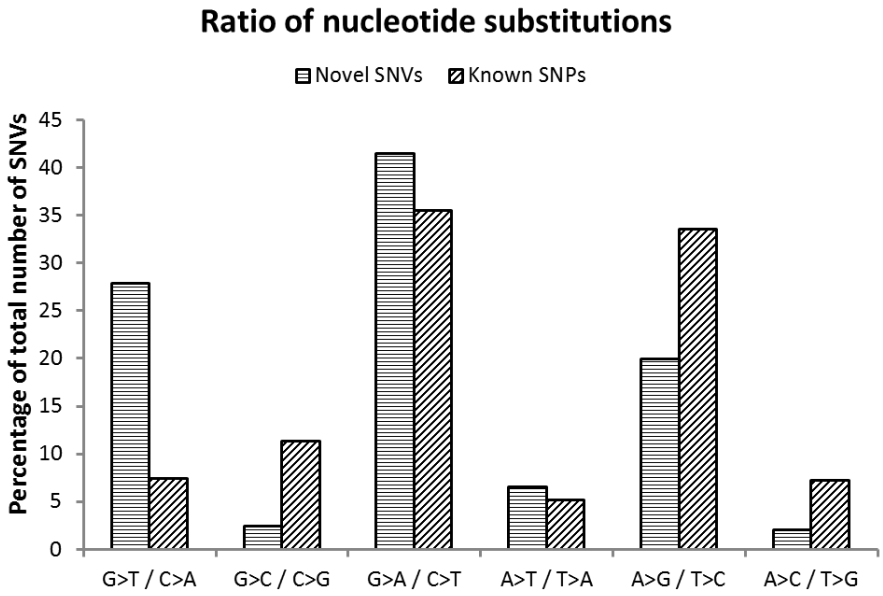



**Figure 3.1.** Mutation spectrum of single nucleotide substitutions. Percentages of mutations in each of the six possible mutation classes.

### *3.4.2. Nucleotide substitution frequencies*

The predominant nucleotide substitutions seen in SNVs of the LNCaP exome are transitions of A <-> G or C <-> T (Figure 3.1). Particularly, we observed that 61.4% of our SNVs are transitions, while 38.6% are transversions, a 1.6:1 ratio. Such a mutation spectrum, which is dominated by C-to-T/G-to-A transitions, has been noted in several other adult cancers such as melanoma, breast, lung, colorectal and ovarian cancers as well as in PCa (Greenman *et al.* 2007, Kumar *et al.* 2011).

Besides the non-synonymous coding variants, other variants were detected as well. Table 3.2 depicts the intragenic locations of the SNVs and the consequence of these variants for the open reading

frames (gain or loss of stop codon, whether or not there is a change in coding sequence). About 22% of all novel variants found after filtering were single-base substitutions that were non-synonymous coding. Only 1.7% of variants result in stop codons, and 8.6% of the variants result in alterations of untranslated regions immediately adjacent to the start and stop codons.

**Table 3.2.** Mutations identified in LNCaP cells

| *Type of change* | *Novel SNVs* | *Known SNPs* |
|---|---|---|
| Non-synonymous coding | 1 802 | 6 823 |
| Synonymous coding | 923 | 7 744 |
| Stop gained | 134 | 63 |
| Stop lost | 4 | 17 |
| Within non-coding gene | 1 702 | 10 244 |
| Intronic | 2 851 | 19 242 |
| 3 prime UTR | 456 | 2 564 |
| 5 prime UTR | 244 | 1 267 |

Abbreviation: UTR, untranslated region

### *3.4.3. Detection of small insertions and deletions*

For the prediction of small (< 7 base pairs) insertions and deletions, we used the Dindel software (Albers *et al.* 2011) resulting in 43767 indels. These indel calls were filtered using a combination of criteria (Table 3.3). With the requirement that the Phred-quality was at least 20, we detected 17155 events of which 8014 (47%) were not previously documented in dbSNP131. Further filtering for indels in coding regions and by strand bias, resulted in 218 indels of which 67 were insertions and 151 deletions. All indels that were found in the LNCaP exome were heterozygous with mutant allele frequencies below 70%. It should be noted that the frameshift in the PTEN gene (Vlietstra *et al.* 1998) was also detected in our data. Figure 3.2 depicts a Circos-plot with an exome-wide representation of all novel homozygous and heterozygous SNVs as well as small insertions and deletions that were found in the exome of the LNCaP cell line.

**Table 3.3.** Filters used to identify indels in the LNCaP exome

| *Filtering criteria* | *Number of indels* |
|---|---|
| Predicted | 43 767 |
| With Phred-quality ≥ 20 | 17 155 |
| Not present in dbSNP131 | 8 014 |
| Change in amino acid sequence | 223 |
| Without strand bias | 218 |

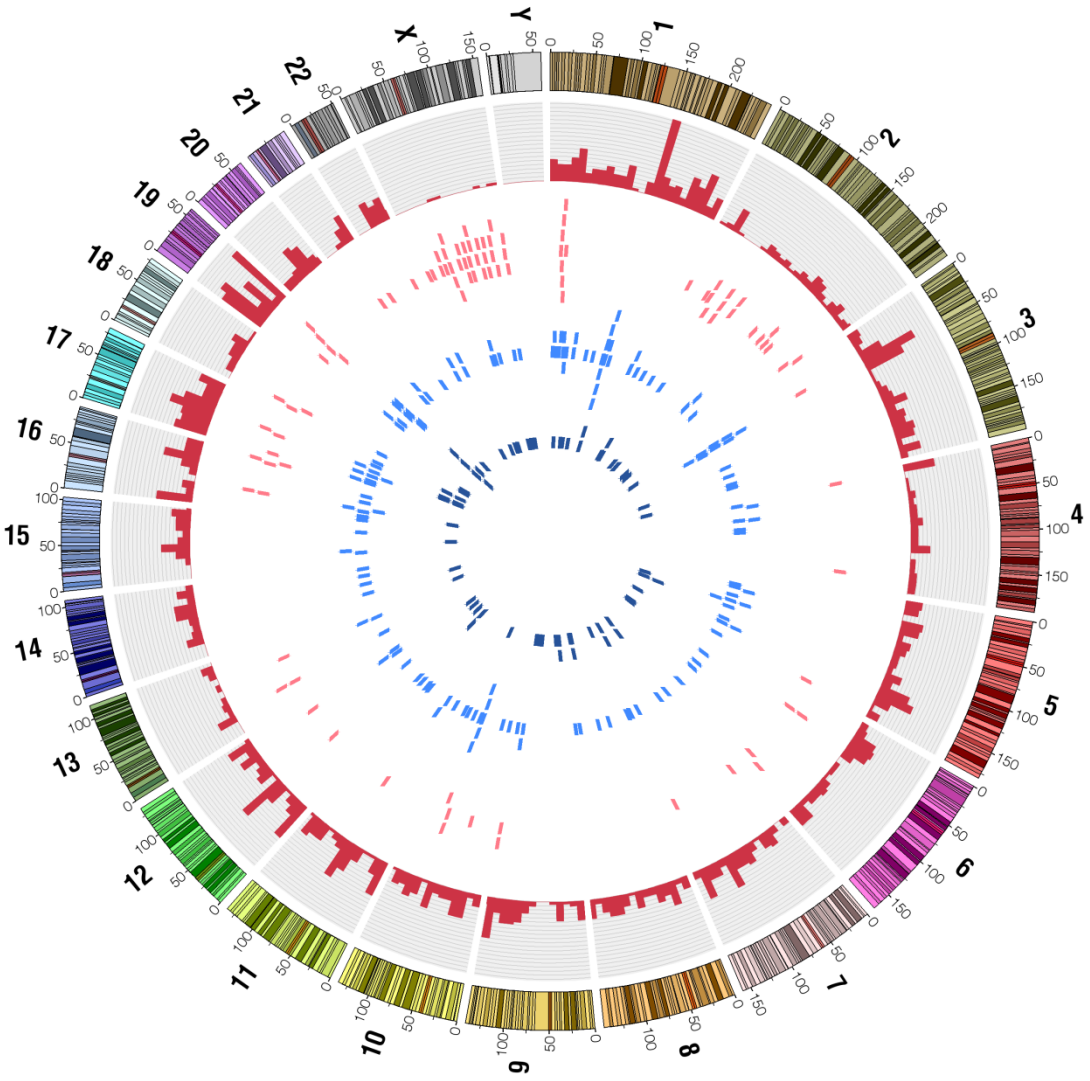


**Figure 3.2.** The catalogue of mutations in LNCaP cells. Figurative representation of the catalogue of mutations in the exome of LNCaP cells. Chromosome ideograms are shown around the outer ring and oriented pter-qter in a clockwise direction with centromeres indicated in red. Other tracks contain alterations (from outside to inside): 1704 heterozygous SNVs shown by density per 10 megabases (dark red bars); 98 homozygous SNVs (light red rectangles), 151 deletions (outer blue rectangles) and 67 insertions (inner blue rectangles).

### *3.4.4. Prediction of impact on protein function*

The high number of identified SNVs obstructs a straightforward analysis of their effects on protein functions and the identification of their role in oncogenesis. However, we wanted to determine the effects of the non-synonymous SNVs identified in our LNCaP cells on the structure and function of the changed proteins. We developed a workflow to calculate a score for the theoretical impact of each SNV.

We first used publicly available software to make *in silico* predictions of the impact that these SNVs could have on protein structure and/or protein function. These predictions were calculated by three different software packages: SIFT, PolyPhen2 and MutationAssessor (Adzhubei *et al.* 2010, Kumar *et al.* 2009, Reva *et al.* 2011). Only 7% of the mutations were predicted by all three methods to have a potential functional effect, whereas 26% of the mutations were predicted to be potentially deleterious by at least one of the three methods. We next focused on those variants for which all three programs predicted an impact and combined these predictions into one ranking (Ranking A in Figure 3.3).

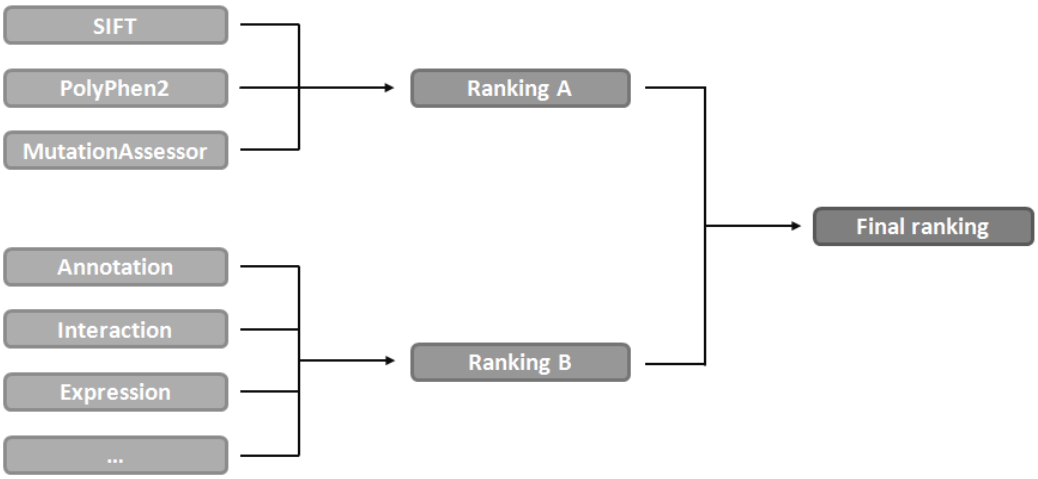



**Figure 3.3.** Prioritization workflow. SIFT, PolyPhen2 and MutationAssessor are *in silico* software programs that predict the potential influence a SNV might have on protein function. All three predictions were combined into ranking A. Gene information such as annotation, protein-protein interaction and expression data was used to generate ranking B. The combination of ranking A and B results in a final ranking that prioritizes the mutated genes and can help in defining genes for future research.

To take into account additional information about the genes in which SNVs are localized, we used the Endeavour software (Tranchevent *et al.* 2008). This approach allowed us to include data on functional and pathway annotation (GO, KEGG), protein-protein interactions, expression levels and literature. Next, the Endeavour-based ranking (Ranking B) was combined with Ranking A to obtain a

functional prioritization of all genes that are mutated in the exome of the LNCaP PCa cell line. A list of the 50 genes that are ranked highest is available as supplementary data (Supplementary Table 3.1). Not surprisingly, nine of the top ten of genes in the final ranking are kinases that could have been activated or inactivated during oncogenesis. From the top 100 SNVs, 12 SNVs are located in genes that have been found to be mutated in PCa genomes (Berger *et al.* 2011b, Kumar *et al.* 2011, Robbins *et al.* 2011, Taylor *et al.* 2010).

### *3.4.5. Validation of SNVs*

To ensure that the observed mutations are no artifacts of the PCR-amplification that is part of the library preparation or of the Illumina sequencing, we independently performed a PCR-based amplification of the exonic regions containing the mutations followed by conventional Sanger sequencing (Figure 3.4). In 115 genes that are reported in literature to be involved in the development of PCa, we detected 14 SNVs and 12 SNPs. All these SNVs and SNPs were validated by capillary sequencing of genomic PCR products (Figure 3.5.A-B). Among these genes is the gene that encodes the androgen receptor (AR) harboring a known mutation in exon 8 (Veldscholte *et al.* 1990). This T877A mutation was detected in all 58 reads covering this base. The expression of all 14 SNV-containing exons was verified by RNA sequencing (Figure 3.5.A). The presence of the 12 known SNPs (Figure 3.5.B) in this group of genes was also confirmed at the genomic DNA level. However, for 4 out of the 12 SNPs (CD44, FLT3, GSTP1 and RNASEL) no transcript could be generated.

We also re-sequenced the regions around each of these SNVs and SNPs in other PCa cell lines (DU145, PC-3, CWR22Rv1 and VCaP). While SNPs could also be detected in these other PCa cell lines, as expected, this was not the case for our new SNVs (Figure 3.5.C-D).

Aside from these SNVs and SNPs, we validated the presence of 16 other SNVs in the genomic DNA as well as in the transcriptome (Supplementary Figure 3.1). In total, for all 42 SNVs that were validated on the genomic LNCaP DNA, the presence of the variant was confirmed with Sanger sequencing. While this is only a small sample of the SNVs, it illustrates the low false positive rate which is a consequence of the stringent filtering method we applied.

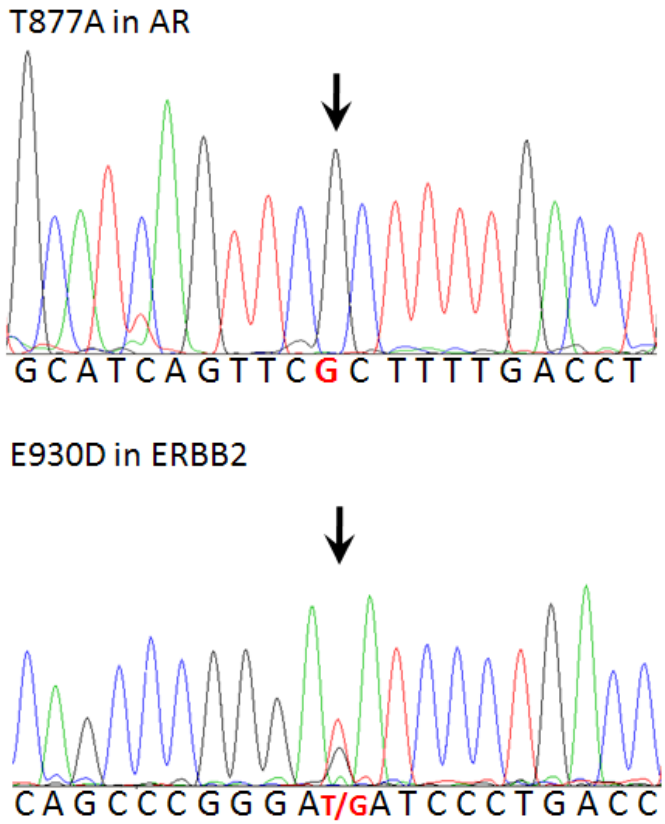



**Figure 3.4.** Validation of SNVs and SNPs with PCR and Sanger sequencing. Sanger sequence traces from non-synonymous mutations in the LNCaP exome. The mutation in the androgen receptor is an example of a homozygous SNV, while the ERBB2 gene contains a heterozygous SNV. The mutated position is arrowed.

### *3.4.6. Genomic stability of the LNCaP cell line*

LNCaP cells are known to show inter-laboratory variability, even when they are all derived from the same tumor sample and most likely even from a single cancer cell. However, it is unclear whether this is due to phenotypic or genotypic adaptations. In addition, the presence of two alternative bases at any specific position could be due to heterozygosity or heterogeneity of the culture. Therefore, the presence of SNVs and SNPs was verified in LNCaPs from different laboratories (LNCaP (UK) and LNCaP (NL)), in a monoclonal cell line derived from LNCaP cells (LNCaP TR2) and in a cell line derived from LNCaP cells after xenografting in nude mice (C4-2B cells). Most SNVs and SNPs were found in all cell lines (Figure 3.5.A-B). This demonstrates that the LNCaP cells are heterozygous for these SNVs and SNPs. Moreover, the fact that the SNVs in the E-cadherin, CDK4, Notch1 and PlexinB1 genes are present in some sublines, but absent in LNCaP TR2 and LNCaP (UK) might indicate some form of genetic instability.

A

| | | LNCaP | cDNA LNCaP | LNCaP (UK) | LNCaP TR2 (UK) | C4-2B | LNCaP (NL) |
|---|---|---|---|---|---|---|---|
| APC | R2714C | + | + | + | + | + | |
| AR | T877A | + | + | + | + | + | |
| ARAF | E280K | + | + | + | + | + | |
| CDH1 | P94T | + | + | - | - | + | + |
| CDK4 | P110L | + | + | - | - | + | + |
| CHEK2 | T378N | + | + | + | + | + | |
| ERBB2 | E930D | + | + | + | + | + | |
| MSH6 | P1073S | + | + | + | + | + | |
| MYC | N30S | + | + | + | + | + | |
| NF1 | A431V | + | + | + | + | + | |
| NOTCH1 | T865N | + | + | - | - | + | + |
| PLXNB1 | T1697A | + | + | - | - | - | - |
| SMO | C273R | + | + | + | + | + | |
| TERT | R865C | + | + | + | + | + | |

C

| | | VCaP | DU145 | PC-3 | CWR 22Rv1 |
|---|---|---|---|---|---|
| APC | R2714C | - | - | - | - |
| AR | T877A | - | - | - | - |
| ARAF | E280K | - | - | - | - |
| CDH1 | P94T | - | - | - | - |
| CDK4 | P110L | - | - | - | - |
| CHEK2 | T378N | - | - | - | - |
| ERBB2 | E930D | - | - | - | - |
| MSH6 | P1073S | - | - | - | - |
| MYC | N30S | - | - | - | - |
| NF1 | A431V | - | - | - | - |
| NOTCH1 | T865N | - | - | - | - |
| PLXNB1 | T1697A | - | - | - | - |
| SMO | C273R | - | - | - | - |
| TERT | R865C | - | - | - | - |

B

| | | LNCaP | cDNA LNCaP | LNCaP (UK) | LNCaP TR2 (UK) | C4-2B |
|---|---|---|---|---|---|---|
| AMACR | V186F | + | + | + | + | + |
| BRCA2 | N372H | + | + | + | + | + |
| CD44 | K417R | + | / | + | + | + |
| CDKN1A | S31R | + | + | + | + | + |
| CYP1B1 | V432L | + | + | + | + | + |
| ERBB2 | P1170A | + | + | + | + | + |
| FLT3 | D358V | + | / | + | + | - |
| GSTP1 | I105V | + | / | + | + | + |
| ITGA7 | R695H | + | + | + | + | + |
| MLH1 | I219V | + | + | + | + | + |
| RNASEL | R462Q | + | / | + | + | + |
| TP53 | P72R | + | + | + | + | + |

D

| | | VCaP | DU145 | PC-3 | CWR 22Rv1 |
|---|---|---|---|---|---|
| AMACR | V186F | - | - | + | - |
| BRCA2 | N372H | - | - | - | - |
| CD44 | K417R | + | - | - | + |
| CDKN1A | S31R | - | - | - | - |
| CYP1B1 | V432L | - | + | + | + |
| ERBB2 | P1170A | + | + | - | + |
| FLT3 | D358V | - | - | - | - |
| GSTP1 | I105V | - | - | + | + |
| ITGA7 | R695H | + | - | - | + |
| MLH1 | I219V | + | - | - | + |
| RNASEL | R462Q | + | - | - | + |
| TP53 | P72R | + | + | - | + |

**Figure 3.5.** Validation of SNVs and SNPs in the LNCaP exome. Validation was performed using PCR on both genomic DNA and copy-DNA, followed by conventional Sanger sequencing. The first and second column of each table represents the name of the gene and the amino acid substitution respectively. **A and B,** validation of SNVs and SNPs respectively on genomic DNA and copy-DNA from the LNCaP cells cultured in our laboratory, from LNCaP cells from other laboratories (LNCaP (UK) and LNCaP (NL)), from a monoclonal derivative of LNCaP cells (LNCaP TR2) and from a cell line derived from LNCaP cells after xenografting in nude mice (C4-2B). **C and D,** validation of SNVs and SNPs respectively on genomic DNA from other PCa cell lines (VCaP, DU145, PC-3 and CWR22Rv1). A + sign denotes that the SNV or SNP was detected, a – sign that the SNV or SNP was not detected and / means that no PCR-product could be amplified.

## 3.5. DISCUSSION

### *3.5.1. Exome analysis*

The re-sequencing of the exome of LNCaP cells generated a comprehensive catalogue of 1802 SNVs, 67 insertions and 151 deletions specific for this PCa cell line. As there is no normal tissue from the corresponding patient available, it is possible that some of the novel variants we found are germ line SNPs not yet reported in the dbSNP database. However, most of them will more likely be mutations that originated either during carcinogenesis or during the culturing of these cells. Our SNV catalogue includes the majority of mutations present in LNCaP cells since it is based on 78% of exonic nucleotides that were covered 12x or more. Extrapolation reveals that, at least in theory, the complete LNCaP exome would contain over 2300 SNVs. With our validation study using PCR-amplification and capillary sequencing, we confirmed the presence of 42 SNVs in the original LNCaP culture, illustrating the quality resulting from our stringent filtering criteria. On the other hand, some SNVs will have escaped detection because we maintained a severe filtering of the sequencing data. A less stringent filtering would lead to the detection of more SNVs, but this would be at the expense of specificity, since false positive SNV predictions would increase. A ten-fold read depth was *e.g.* applied by Chang *et al.*, but the rate of false positive SNVs was not reported (Chang *et al.* 2011).

At first, we looked for SNVs and SNPs in 115 cancer-related genes that had been connected with PCa in literature before. From these, we found 25 genes to contain a SNV and/or SNP in the LNCaP exome (Figure 3.5). For twenty-two of the 25 genes, other mutations had already been described in PCa biopsies (Assinder *et al.* 2009, Dong 2006, Grindedal *et al.* 2009, Hughes *et al.* 2005, Mimeault *et al.* 2006, Wong *et al.* 2007). Furthermore, 16 of these genes were found to be mutated in a limited number of PCa samples analyzed with next-generation sequencing technologies (Berger *et al.* 2011b, Kumar *et al.* 2011, Robbins *et al.* 2011, Taylor *et al.* 2010).

For most tested SNVs, expression was demonstrated by reverse transcriptase-PCR of LNCaP polyA-RNA. Even when two alternative bases were called in the high-throughput sequencing, both could be detected in the RNA, and when a homozygous SNV or SNP was validated, only this base was detected in the transcriptome. Thus, for all SNVs both the wild type and the mutated genes are expressed in the LNCaP cells. However, for four SNPs, we could not detect transcripts, and this is in accordance with microarray data available for LNCaP cells indicating that these genes are not expressed in LNCaP cells (Oudes *et al.* 2005). However, we cannot conclude whether the repression of these variant alleles happened before, during or even after the oncogenic process. Not surprisingly, none of the SNVs and only part of the SNPs that were tested (Figure 3.5) are present in other PCa cell lines (VCaP, DU145, PC-3 and CWR22Rv1).

The number of SNVs identified in the LNCaP cell line is higher than what we would have predicted, certainly when compared to the number of SNVs found in the genome or exome of other cancer cell lines, even when a less stringent filtering was applied (Chang *et al.* 2011, Clark *et al.* 2010). For example, whole genome sequencing of the U87MG glioma cell line identified 1036 non-synonymous coding SNVs with less strict filtering criteria, while whole exome sequencing of the PC-3 PCa cell line identified only 200 high-confidence non-synonymous coding SNVs (Chang *et al.* 2011, Clark *et al.* 2010).

### *3.5.2. Defining prostate cancer driver mutations*

Prostate adenocarcinoma is the most frequently diagnosed malignancy in men and it is the second leading cause of cancer deaths among men in Western countries (Parkin *et al.* 2005). Its biological heterogeneity is a major cause of the overtreatment and undertreatment of 30 to 50% of PCa patients (Cooperberg *et al.* 2005). The limited genomic data available from PCa biopsies indicate that PCa seems to have a low frequency of mutations (Kan *et al.* 2010, Taylor *et al.* 2010). The most frequent PCa-related mutations are gene rearrangements (Tomlins *et al.* 2005), but these were not tested in this study.

Identifying genomic changes in PCa would be a valuable diagnostic tool (Berger *et al.* 2011b, Robbins *et al.* 2011, Taylor *et al.* 2010), but as for all exome and genome sequencing studies, distinguishing which mutations matter is challenging. Indeed, cancer somatic mutations can be subdivided in two main biological classes: driver mutations, which confer selective clonal growth advantages, are causally implicated in oncogenesis, while passenger mutations have not contributed to the development of the cancer and hence have not been subjected to direct selection. It should be noted here that when looking at variants in cell lines, these could have originated even after the carcinogenesis during the process of *in vitro* culturing of the cells.

The ratio of non-synonymous to synonymous mutations is a first reliable indicator for some form of selective pressure, as synonymous alterations are unlikely to exert a selective growth advantage (Bardelli *et al.* 2003). Applied to our data, we found a ratio of 2.1:1 which is not higher than the ratio of 2:1 predicted for non-selected passenger mutations. This information, combined with the fact that only 7% of all SNVs are predicted to be deleterious according to all three *in silico* programs (SIFT, PolyPhen2 and MutationAssessor), suggests that the majority of the SNVs we identified are functionally neutral and most likely ‘passenger’ mutations. It should be noted however that the classical discrete, step-wise model of tumor suppression is being replaced by the concept of a dosage-dependency of tumor suppressor gene function influenced by smaller changes in many gene

products (Berger *et al.* 2011a). In this concept, many of the variants that would be called passenger mutations based on the Endeavour protocol could at some point have contributed to the oncogenic process.

In the present study, we focused on non-synonymous coding variants. However, it has been suggested that synonymous variants or variants in untranslated regions could still have a powerful impact by altering the level of gene expression through alteration of miRNA binding (Berger *et al.* 2011a).

### *3.5.3. In silico search for driver mutations*

Our exome re-sequencing data set was used to devise a prioritization protocol which returns a list of SNVs in genes that might have an impact on protein structure or function. For Ranking A, we choose to combine three separate prediction software programs, because each program has its own assumptions and its own advantages and disadvantages. Ranking B was created using the Endeavour software and includes different gene information sources. Furthermore, it allowed the combination of ranking A and B. The observation that nine of the top ten of genes in the final ranking are kinases has implications for the use of LNCaP cells as a model for PCa, for example in preclinical testing of kinase inhibitors or in the study of gene knock down by siRNA. A gene ontology analysis of the top hundred of highest ranked genes resulted in the following most perturbed cell functions: cell death, cellular growth and proliferation, cellular development, cell cycle and cellular movement.

### *3.5.4. Interference with androgen responsiveness*

Initial androgen responsiveness and later acquired castration resistance are important topics in PCa research. First of all, we confirmed that the LNCaP cell line is homozygous for the 877 mutation in the AR gene. In addition, some of our highly ranked proteins (CDK9 and HoxB13) are known to be involved in the androgen response (Gordon *et al.* 2010, Norris *et al.* 2009). Moreover, protein kinase C beta, recently described as an important player in the androgen-induced histone code (Metzger *et al.* 2010), is heterozygous for K489 to N in LNCaP cells. Taylor *et al.* reported a higher frequency of mutations in the AR pathway (including several known AR coactivators and corepressors). We detected 21 mutations in possible AR coregulators, but these need further analysis to determine their impact on the androgen response. Indeed, whether these LNCaP mutations or combinations thereof have contributed to the carcinogenesis, the metastasis or the evolution into castration-resistance as commonly seen for PCa remains to be determined.

### *3.5.5. Are LNCaP cells genetically unstable, heterogeneous and/or heterozygous?*

LNCaP cells are homozygous for only 98 of the 1802 SNVs, which is in agreement with the reported polyploidy of LNCaP cells and a limited loss of heterozygosity (Horoszewicz *et al.* 1983). For most SNVs (94.6%), we detected two alternative nucleotides. This could be the consequence of either heterozygosity or heterogeneity of the LNCaP cells. However, most SNVs and SNPs were also present in genomic DNA of LNCaP cells from another source (LNCaP UK), as well as from a monoclonal derivative (LNCaP TR2) and in the syngeneic C4-2B cell line (Figure 3.5). Since heterogeneity would most likely have been lost during the monoclonal derivation, the presence of two alternatives at most SNVs can only be explained by the fact that the LNCaP cells are heterozygous rather than heterogeneous for these SNVs and SNPs.

Surprisingly, four out of 14 SNVs were absent in the LNCaP (UK) cells as well as in the derived LNCaP TR2 clone. The presence of three of these four SNVs (E-cadherin, CDK4, and Notch1) was confirmed in LNCaP cells from a third source (LNCaP NL) as well as in the C4-2B cell line. The fact that some LNCaP cell lines have lost heterozygosity of some SNVs indicates that the LNCaP genome is unstable to some extent. Interestingly, the SNV in the PlexinB1 gene was only found in our LNCaP cells and not in any of the other cell lines we tested. This mutation was already reported in LNCaP cells (Wong *et al.* 2007) and its presence was proposed to be linked with an adaptation of the cells to the growth conditions on plastic surfaces, as the mutant protein had significantly increased adhesion capacities.

## 3.6. CONCLUSIONS

In conclusion, we present a nearly complete exome sequence of the LNCaP PCa cell line. We report 1802 non-synonymous single nucleotide variants, 67 small insertions and 151 small deletions. The presence of SNVs detected in our LNCaP cells in other LNCaP (derivative) cell lines indicates that they are mainly heterozygous, but to some degree also heterogeneous or genomically unstable. Although this is accepted for most cancer cell lines, this remained to be documented for LNCaP cells. The SNVs, the SNPs and the heterogeneity between cultures should be taken into account when performing experiments in LNCaP cells for basic as well as preclinical studies.

### 3.7. REFERENCES


Adzhubei, I. A., Schmidt, S., Peshkin, L*., et al.* (2010). "A method and server for predicting damaging missense mutations." Nat Methods **7**(4): 248-249.

Albers, C. A., Lunter, G., MacArthur, D. G*., et al.* (2011). "Dindel: accurate indel calls from short-read data." Genome Res **21**(6): 961-973.

Assinder, S. J., Dong, Q., Kovacevic, Z*., et al.* (2009). "The TGF-beta, PI3K/Akt and PTEN pathways: established and proposed biochemical integration in prostate cancer." Biochem J **417**(2): 411-421.

Bainbridge, M. N., Warren, R. L., Hirst, M*., et al.* (2006). "Analysis of the prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach." BMC Genomics **7**: 246.

Bardelli, A., Parsons, D. W., Silliman, N*., et al.* (2003). "Mutational analysis of the tyrosine kinome in colorectal cancers." Science **300**(5621): 949.

Berger, A. H., Knudson, A. G. and Pandolfi, P. P. (2011a). "A continuum model for tumour suppression." Nature **476**(7359): 163-169.

Berger, M. F., Lawrence, M. S., Demichelis, F*., et al.* (2011b). "The genomic complexity of primary human prostate cancer." Nature **470**(7333): 214-220.

Blankenberg, D., Von Kuster, G., Coraor, N*., et al.* (2010). "Galaxy: a web-based genome analysis tool for experimentalists." Curr Protoc Mol Biol **Chapter 19**: Unit 19 10 11-21.

Chang, H., Jackson, D. G., Kayne, P. S*., et al.* (2011). "Exome sequencing reveals comprehensive genomic alterations across eight cancer cell lines." PLoS One **6**(6): e21097.

Clark, M. J., Homer, N., O'Connor, B. D*., et al.* (2010). "U87MG decoded: the genomic sequence of a cytogenetically aberrant human cancer cell line." PLoS Genet **6**(1): e1000832.

Cooperberg, M. R., Moul, J. W. and Carroll, P. R. (2005). "The changing face of prostate cancer." J Clin Oncol **23**(32): 8146-8151.

Dong, J. T. (2006). "Prevalent mutations in prostate cancer." J Cell Biochem **97**(3): 433-447.

Esquenet, M., Swinnen, J. V., Heyns, W*., et al.* (1997). "LNCaP prostatic adenocarcinoma cells derived from low and high passage numbers display divergent responses not only to androgens but also to retinoids." J Steroid Biochem Mol Biol **62**(5-6): 391-399.

Goecks, J., Nekrutenko, A., Taylor, J*., et al.* (2010). "Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences." Genome Biol **11**(8): R86.

Gordon, V., Bhadel, S., Wunderlich, W*., et al.* (2010). "CDK9 regulates AR promoter selectivity and cell growth through serine 81 phosphorylation." Mol Endocrinol **24**(12): 2267-2280.

Greenman, C., Stephens, P., Smith, R*., et al.* (2007). "Patterns of somatic mutation in human cancer genomes." Nature **446**(7132): 153-158.

Grindedal, E. M., Moller, P., Eeles, R*., et al.* (2009). "Germ-line mutations in mismatch repair genes associated with prostate cancer." Cancer Epidemiol Biomarkers Prev **18**(9): 2460-2467.

Horoszewicz, J. S., Leong, S. S., Chu, T. M*., et al.* (1980). "The LNCaP cell line--a new model for studies on human prostatic carcinoma." Prog Clin Biol Res **37**: 115-132.

Horoszewicz, J. S., Leong, S. S., Kawinski, E*., et al.* (1983). "LNCaP model of human prostatic carcinoma." Cancer Res **43**(4): 1809-1818.

Hughes, C., Murphy, A., Martin, C*., et al.* (2005). "Molecular pathology of prostate cancer." J Clin Pathol **58**(7): 673-684.

Kan, Z., Jaiswal, B. S., Stinson, J*., et al.* (2010). "Diverse somatic mutation patterns and pathway alterations in human cancers." Nature **466**(7308): 869-873.

Koboldt, D. C., Chen, K., Wylie, T*., et al.* (2009). "VarScan: variant detection in massively parallel sequencing of individual and pooled samples." Bioinformatics **25**(17): 2283-2285.

Krzywinski, M., Schein, J., Birol, I*., et al.* (2009). "Circos: an information aesthetic for comparative genomics." Genome Res **19**(9): 1639-1645.

Kumar, A., White, T. A., Mackenzie, A. P.*, et al.* (2011). "Exome sequencing identifies a spectrum of mutation frequencies in advanced and lethal prostate cancers." Proc Natl Acad Sci U S A **108**(41): 17087-17092.

Kumar, P., Henikoff, S. and Ng, P. C. (2009). "Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm." Nat Protoc **4**(7): 1073-1081.

Langmead, B., Trapnell, C., Pop, M.*, et al.* (2009). "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome." Genome Biol **10**(3): R25.

Li, H. and Durbin, R. (2009a). "Fast and accurate short read alignment with Burrows-Wheeler transform." Bioinformatics **25**(14): 1754-1760.

Li, H., Handsaker, B., Wysoker, A.*, et al.* (2009b). "The Sequence Alignment/Map format and SAMtools." Bioinformatics **25**(16): 2078-2079.

Metzger, E., Imhof, A., Patel, D.*, et al.* (2010). "Phosphorylation of histone H3T6 by PKCbeta(I) controls demethylation at histone H3K4." Nature **464**(7289): 792-796.

Mimeault, M. and Batra, S. K. (2006). "Recent advances on multiple tumorigenic cascades involved in prostatic cancer progression and targeting therapies." Carcinogenesis **27**(1): 1-22.

Norris, J. D., Chang, C. Y., Wittmann, B. M.*, et al.* (2009). "The homeodomain protein HOXB13 regulates the cellular response to androgens." Mol Cell **36**(3): 405-416.

Oudes, A. J., Roach, J. C., Walashek, L. S.*, et al.* (2005). "Application of Affymetrix array and Massively Parallel Signature Sequencing for identification of genes involved in prostate cancer progression." BMC Cancer **5**: 86.

Pan, Y., Kytola, S., Farnebo, F.*, et al.* (1999). "Characterization of chromosomal abnormalities in prostate cancer cell lines by spectral karyotyping." Cytogenet Cell Genet **87**(3-4): 225-232.

Parkin, D. M., Bray, F., Ferlay, J.*, et al.* (2005). "Global cancer statistics, 2002." CA Cancer J Clin **55**(2): 74-108.

Quinlan, A. R. and Hall, I. M. (2010). "BEDTools: a flexible suite of utilities for comparing genomic features." Bioinformatics **26**(6): 841-842.

Reva, B., Antipin, Y. and Sander, C. (2011). "Predicting the functional impact of protein mutations: application to cancer genomics." Nucleic Acids Res **39**(17): e118.

Robbins, C. M., Tembe, W. A., Baker, A.*, et al.* (2011). "Copy number and targeted mutational analysis reveals novel somatic events in metastatic prostate tumors." Genome Res **21**(1): 47-55.

Robinson, J. T., Thorvaldsdottir, H., Winckler, W.*, et al.* (2011). "Integrative genomics viewer." Nat Biotechnol **29**(1): 24-26.

Taylor, B. S., Schultz, N., Hieronymus, H.*, et al.* (2010). "Integrative genomic profiling of human prostate cancer." Cancer Cell **18**(1): 11-22.

Tomlins, S. A., Rhodes, D. R., Perner, S.*, et al.* (2005). "Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer." Science **310**(5748): 644-648.

Tranchevent, L. C., Barriot, R., Yu, S.*, et al.* (2008). "ENDEAVOUR update: a web resource for gene prioritization in multiple species." Nucleic Acids Res **36**(Web Server issue): W377-384.

van Bokhoven, A., Varella-Garcia, M., Korch, C.*, et al.* (2003). "Molecular characterization of human prostate carcinoma cell lines." Prostate **57**(3): 205-225.

Veldscholte, J., Ris-Stalpers, C., Kuiper, G. G.*, et al.* (1990). "A mutation in the ligand binding domain of the androgen receptor of human LNCaP cells affects steroid binding characteristics and response to anti-androgens." Biochem Biophys Res Commun **173**(2): 534-540.

Vlietstra, R. J., van Alewijk, D. C., Hermans, K. G.*, et al.* (1998). "Frequent inactivation of PTEN in prostate cancer cell lines and xenografts." Cancer Res **58**(13): 2720-2723.

Wong, O. G., Nitkunan, T., Oinuma, I.*, et al.* (2007). "Plexin-B1 mutations in prostate cancer." Proc Natl Acad Sci U S A **104**(48): 19040-19045.

### 3.8. SUPPLEMENTARY INFORMATION

**Supplementary table 3.1.** Top 50 of genes ranked highly with *in silico* prioritization protocol. This table lists 50 genes that were ranked highest using our *in silico* prioritization protocol.

| | | |
|---|---|---|
| 1 | RIPK2 | Receptor-interacting serine/threonine-protein kinase 2 |
| 2 | CHEK2 | Serine/threonine-protein kinase Chk2 |
| 3 | PRKCB1 | Protein kinase C beta type |
| 4 | RET | Proto-oncogene tyrosine-protein kinase receptor ret precursor |
| 5 | CSNK1G2 | Casein kinase I isoform gamma-2 |
| 6 | MELK | Maternal embryonic leucine zipper kinase |
| 7 | ZAP70 | Tyrosine-protein kinase ZAP-70 |
| 8 | CDK9 | Cell division protein kinase 9 (Cyclin-dependent kinase 9) |
| 9 | KLK3 | Prostate-specific antigen precursor (Kallikrein- 3) |
| 10 | GAB1 | GRB2-associated-binding protein 1 |
| 11 | FLT4 | Vascular endothelial growth factor receptor 3 precursor |
| 12 | COL5A1 | Collagen alpha-1(V) chain precursor |
| 13 | TERT | Telomerase reverse transcriptase |
| 14 | PIK3R4 | Phosphoinositide 3-kinase regulatory subunit 4 |
| 15 | HSD17B4 | 17-beta-hydroxysteroid dehydrogenase 4 |
| 16 | EHMT1 | Euchromatic Histone-lysine N-methyltransferase 1 |
| 17 | PTK7 | Tyrosine-protein kinase-like 7 precursor |
| 18 | CAMK2G | Calcium/calmodulin-dependent protein kinase type II gamma chain |
| 19 | SMO | Smoothened homolog precursor |
| 20 | DAB2 | Disabled homolog 2 |
| 21 | CAPZA1 | F-actin capping protein subunit alpha-1 |
| 22 | ATP1A1 | Sodium/potassium-transporting ATPase alpha-1 chain precursor |
| 23 | CAMK2B | Calcium/calmodulin-dependent protein kinase type II beta chain |
| 24 | TLN1 | Talin-1 |
| 25 | MYC | Myc proto-oncogene protein |
| 26 | CBLB | E3 ubiquitin-protein ligase CBL-B |
| 27 | SOCS7 | Suppressor of cytokine signaling 7 |
| 28 | RASGRP2 | RAS guanyl releasing protein 2 isoform 2 |
| 29 | MYLK | Myosin light chain kinase, smooth muscle |
| 30 | HNRPH3 | Heterogeneous nuclear ribonucleoprotein H3 |
| 31 | VAV1 | Proto-oncogene vav |
| 32 | GNA11 | Guanine nucleotide-binding protein subunit alpha-11 |
| 33 | GNAI3 | Guanine nucleotide-binding protein G(k) subunit alpha-3 |
| 34 | UCP2 | Mitochondrial uncoupling protein 2 |
| 35 | RBL2 | Retinoblastoma-like protein 2 |
| 36 | NSMAF | Protein FAN (Factor associated with neutral sphingomyelinase activation) |
| 37 | LTA4H | Leukotriene A-4 hydrolase |
| 38 | MYH14 | Myosin heavy chain 14 |
| 39 | DIABLO | Diablo homolog, mitochondrial precursor |
| 40 | ABI1 | Abl interactor 1 (Abelson interactor 1) |

| | | |
|---|---|---|
| 41 | ELAVL1 | ELAV-like protein 1 |
| 42 | G3BP1 | Ras GTPase-activating protein-binding protein 1 |
| 43 | HSD17B10 | 3-hydroxyacyl-CoA dehydrogenase type-2 |
| 44 | TCEA2 | Transcription elongation factor A protein 2 |
| 45 | FGF5 | Fibroblast growth factor 5 precursor |
| 46 | AGA | N(4)-(beta-N-acetylglucosaminyl)-L-asparaginase precursor |
| 47 | EPHB4 | Ephrin type-B receptor 4 precursor |
| 48 | TSC22D3 | TSC22 domain family protein 3 |
| 49 | RNF34 | E3 ubiquitin-protein ligase RNF34 (RING finger protein 34) |
| 50 | HNRPL | Heterogeneous nuclear ribonucleoprotein L |

| | | LNCaP | cDNA LNCaP |
|---|---|---|---|
| RIPK2 | H144Q | + | + |
| CHEK2 | T387N | + | + |
| PRKCB1 | K489N | + | - |
| RET | R77H | + | + |
| RET | I858T | + | + |
| CSNK1G2 | T215A | + | + |
| MELK | A119D | + | + |
| ZAP70 | R353C | + | / |
| ZAP70 | L371Q | + | / |
| CDK9 | D290Y | + | + |
| KLK3 | P97T | + | + |
| KLK3 | V136M | + | + |
| KLK3 | L218P | + | + |
| GAB1 | C115Y | + | + |
| HoxB13 | L144P | + | + |
| TRIM24 | L328M | + | + |

**Supplementary figure 3.1.** Validation of additional SNVs in the LNCaP exome. The first and second column of each table represents the name of the gene and the amino acid substitution respectively. The validation was performed using PCR on both genomic DNA (third column) and copy-DNA (fourth column), followed by conventional Sanger sequencing. A + sign denotes that the SNV or SNP was detected, a – sign that the SNV or SNP was not detected and / means that no PCR-product could be amplified.

**CHAPTER 4:**

# COMPARATIVE GENOMIC AND TRANSCRIPTOMIC ANALYSES OF LNCAP AND C4-2B PROSTATE CANCER CELL LINES

Lien Spans, Christine Helsen, Liesbeth Clinckemalie, Thomas Van den Broeck, Stefan Prekovic, Steven Joniau, Evelyne Lerut and Frank Claessens

PLoS One; **9**(2); e90002; 2014

The work presented in this chapter has been conducted by the PhD candidate.

## 4.1. ABSTRACT

The LNCaP and C4-2B cell lines form an excellent preclinical model to study the development of metastatic castration-resistant prostate cancer, since C4-2B cells were derived from a bone metastasis that grew in nude mice after inoculation with the LNCaP-derived, castration-resistant C4-2 cells. Exome sequencing detected 2188 and 3840 mutations in LNCaP and C4-2B cells respectively, of which 1784 were found in both cell lines. Surprisingly, the parental LNCaP cells have over 400 mutations that were not found in the C4-2B genome. More than half of the mutations found in the exomes were confirmed by analyzing the RNA-seq data and we observed that the expressed genes are more prone to mutations than non-expressed genes. The transcriptomes also revealed that 457 genes show increased expression and 246 genes show decreased expression in C4-2B compared to LNCaP cells.

By combining the list of C4-2B-specific mutations with the list of differentially expressed genes, we detected important changes in the focal adhesion and ECM-receptor interaction pathways. Integration of these pathways converges on the myosin light chain kinase gene (MLCK) which might contribute to the metastatic potential of C4-2B cells.

In conclusion, we provide extensive databases for mutated genes and differentially expressed genes in the LNCaP and C4-2B prostate cancer cell lines. These can be useful for other researchers using these cell models.

## 4.2. INTRODUCTION

Prostate cancer (PCa) is the most frequently diagnosed cancer and third leading cause of death amongst men in Europe (Ferlay *et al.* 2013). Despite its prevalence, a majority of men is diagnosed with localized, low-risk PCa and would never die because of their cancer when left untreated (Lu-Yao *et al.* 2009). However, patients with high-risk and especially metastatic disease have a much higher risk of dying from PCa with reported PCa-specific mortality rates up to 28.8% for high-risk disease and 66.1% for metastatic disease at 10-years follow-up (Rider *et al.* 2013). Recent epidemiological data have shown that almost 10% of all PCa patients are metastatic at the time of diagnosis, underlining the clinical importance of developing a better insight in the underlying mechanisms of metastatic PCa (Siegel *et al.* 2012). The genomic and transcriptomic changes that accompany the transformation of localized disease to metastatic castration-resistant PCa are being discovered, but are obstructed by the difficulties to obtain biopsies from the different stages of the disease (Haffner *et al.* 2013, Spans *et al.* 2013).

As an alternative, cell lines can be used as models to study the transition to metastatic castration-resistant PCa (Sampson *et al.* 2013). One of the best studied PCa cell lines undoubtedly is the LNCaP cell line. This cell line was derived from a needle biopsy taken from the left supraclavicular lymph node of a 50-year old Caucasian male (Horoszewicz *et al.* 1980). This patient suffered from a rapidly progressing PCa with minimal and brief response to hormonal therapy and no response to chemotherapy. Subsequently, the C4-2 subline was derived from a tumor that developed in castrated nude mice injected with LNCaP cells. Finally, the C4-2B cell line was derived from a bone metastasis after orthotopic transplantation of C4-2 cells in nude mice (Ianculescu *et al.* 2012, Thalmann *et al.* 1994). In other words, C4-2B is a metastatic derivative of the LNCaP cells. The LNCaP and C4-2B progression model therefore mimics the disease advancing from poorly tumorigenic, androgen-sensitive and non-metastatic in LNCaP, to metastatic and androgen-insensitive (or castration-resistant) in C4-2B.

For these two cell lines, changes in karyotype and genomic copy numbers, some point mutations, insertions and deletions have been described, but the comparison of the exome sequences has not been reported yet (Pan *et al.* 1999, Thalmann *et al.* 1994). The first goal of this study was therefore to obtain comprehensive exome data for LNCaP and C4-2B cells. Of course, a comparison of these mutational landscapes only makes sense in the presence of information on the activity of the affected genes. The latter was obtained from transcriptome analyses.

A first step to catalogue point mutations, insertions and deletions in the LNCaP cells was reported in Spans *et al.* (Spans *et al.* 2012). Here, we report on a comparative whole exome and transcriptome

sequencing study of both LNCaP and C4-2B cell lines. To our knowledge, this is the first direct and thorough comparison of this kind. Moreover, these databases can be very informative for preclinical studies for which both LNCaP and C4-2B cells are being used. They can also be used to generate hypotheses on the metastatic process, as exemplified for the MLCK pathway.

## 4.3. MATERIALS AND METHODS

### *4.3.1. DNA isolation*

The LNCaP cell line was obtained from the American Type Culture Collection, while the C4-2B cells were a kind gift from Dr. M. Stallcup (Norris Comprehensive Cancer Center, University of Southern California, USA) (Thalmann *et al.* 1994). Both cell lines were grown in Roswell Park Memorial Institute medium (RPMI, Gibco, Invitrogen), containing 2 g/L glucose supplemented with 10% heat-inactivated fetal calf serum (FCS). The passage number of the LNCaP and C4-2B cells was 48 and 42 respectively. High-molecular weight DNA was extracted from cultured cells using the GenElute Mammalian Genomic DNA Miniprep kit (Sigma-Aldrich). After purification using ethanol precipitation with ammonium acetate, the concentration was quantified using a Nanodrop ND-1000 spectrophotometer (Thermo Fisher Scientific) and BioAnalyzer (Agilent).

### *4.3.2. Whole exome sequencing*

Whole exome capture of the LNCaP cells was performed using the SureSelect Human All Exon System (Agilent) according to the manufacturer's instructions. Paired-end, 100 bp long sequencing reads were generated using the GAIIx sequencer (Illumina). The exome capture of the C4-2B cells was performed using the SeqCap EZ Exome version 2 kit (Roche Nimblegen) and paired-end 100 bp long reads were generated using the HiSeq2000 (Illumina).

Quality control was performed using FastQC software (version 0.10.1) (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/) and Picard (version 1.22) (http://picard.sourceforge.net/). Sequencing reads were aligned to the human reference genome (hg19, NCBI Build 37) using BWA, where reads were trimmed when the quality was below 15 (version 0.5.9) (Li *et al.* 2009a). Alignment files were processed further with Genome Analysis Toolkit (GATK) before variant calling and included duplicate removal, local realignment around known indels and base quality recalibration (version 1.0.5777) (McKenna *et al.* 2010). The samples were loaded individually to the GATK UnifiedGenotyper software. Point mutations and expression data were

plotted using the Circos software (version 0.52) (Krzywinski *et al.* 2009). Comparison of point mutations was performed using Venny (http://bioinfogp.cnb.csic.es/tools/venny/index.html).

### *4.3.3. RNA isolation*

LNCaP and C4-2B cells, with passage numbers of 30 and 43 respectively, were plated in 6-well plates (1.75 million cells/well) and treated overnight (18h) with 1 nM R1881 (Perkin Elmer). The cells were collected and washed with PBS. The cell pellet was used to extract total RNA using the RNeasy Mini Kit from Qiagen. The quality and purity of the RNA was inspected on a Nanodrop ND-1000 Spectrophotometer. The integrity of the RNA was verified on the BioAnalyzer at the Genomics Core of UZ Leuven.

### *4.3.4. RNA sequencing*

After selection of polyA+ RNA, the RNA was converted into cDNA libraries using the TruSeq RNA Sample Preparation kit (Illumina). After sequencing paired-end short reads of 100 bp with the HiSeq2000 (Illumina), normalized gene counts (Fragments Per Kilobase per Million of mapped reads, FPKM) were calculated via the Tuxedo pipeline (Tophat – Cufflinks – CummeRbund) (Trapnell *et al.* 2012). In short, the RNA-seq data were aligned to the reference genome using TopHat (version 2.0.6) that utilizes Bowtie as the algorithmic core. The Cufflinks package (version 2.0.2) assembled the transcripts and detected differentially expressed genes and transcripts. CummeRbund (version 2.0.0) was used to visualize the gene expression data. Variant calling using the RNA-seq data was performed with GATK (version 2.2), after alignment with Tophat (McKenna *et al.* 2010). RNA-seq for both cell lines was performed in triplicate, allowing the identification of differentially expressed genes. For variant calling, the triplicates were aggregated to obtain higher coverage. Pathway-Express was used to determine, from a list of genes, whether in a specific pathway more genes are involved than would be expected by chance (Draghici *et al.* 2007).

### *4.3.5. Quantitative RT-PCR*

cDNA was generated from RNA (1 µg) with Random Hexamer primers and RevertAid Reverse Transcriptase (Thermo Scientific). Quantitative Real Time PCR was performed using Platinum SYBR Green QPCR Supermix-UDG (Invitrogen). Results were normalized to the housekeeping gene beta-actin and each sample was analyzed in triplicate. The sequence of the primers used are: beta-actin forward 5'-ACCCAAGGCCAACCG-3' and reverse 5'-TGTACGTTGCTATCCAGGCTGT-3', TMPRSS2

forward 5'-CCTGCATCAACCCCTCTAACTG-3' and reverse 5'-AGGCGAACACACCGATTCTC-3', IGF1 forward 5'-TGGATGCTCTTCAGTTCGTG-3' and reverse 5'-TCATCCACGATGCCTGTCT-3', IGF1R forward 5'-GTACAACTACCGCTGCTGGA-3' and reverse 5'-TGGCAGCACTCATTGTTCTC-3'.

#### *4.3.6. Accession numbers*

Binary sequence alignment/map (BAM) files from whole exome sequencing data as well as RNA-seq data were deposited in the database of the European Nucleotide Archive with accession number PRJEB4877 and are accessible via http://www.ebi.ac.uk/ena/data/view/PRJEB4877. The sample accession numbers are ERS363578 and ERS363580 for whole exome sequencing data of LNCaP and C4-2B respectively. For the RNA-sequencing, the sample accession numbers are ERS363579 and ERS363581 for LNCaP and C4-2B cells respectively.

#### *4.3.7. Confirmation of non-synonymous variants*

Variants of interest were confirmed by Sanger sequencing of amplified PCR products. Primers specific to the region containing the variant to be tested were designed using the NCBI Primer-Blast (http://ncbi.nlm.nih.gov/tools/primer-blast/) and obtained from Integrated DNA Technologies. Polymerase chain reactions were performed following standard protocols using Taq DNA polymerase (Thermo Scientific). Amplification of specific PCR fragments was confirmed by agarose gel electrophoresis. Sanger sequencing was performed at LGC Genomics. Sequence trace files were analyzed using Chromas Lite.

### 4.4. RESULTS

#### *4.4.1. Detecting point mutations with whole exome sequencing*

We performed a whole-exome re-sequencing study for both LNCaP and C4-2B cells using 100 base pair, paired-end reads on the Illumina platform. This generated 49 and 80 million reads for LNCaP and C4-2B respectively (Supplementary table 4.1); for LNCaP cells, 74% of the exome was covered at least 20x, versus 88% for C4-2B cells. Sequencing characteristics and quality control data are similar for both datasets (Supplementary table 4.1 and Supplementary figure 4.1).

The point mutations in the exomes were detected using the GATK pipeline to which additional filtering was applied: only mutations which had at least 12x coverage and a mutation frequency

above 30% were taken into account. Data were also filtered for absence of the base pair change in dbSNP130. Furthermore, strand bias was eliminated manually (Supplementary table 4.2). This resulted in lists of 2188 and 3840 non-synonymous point mutations in LNCaP and C4-2B cells, respectively (Supplementary table 4.3 and 4.4). Only 1784 mutations were common between both cell lines, clearly indicating the accumulation of more than 2000 additional mutations in the C4-2B genome. This large difference in mutation load cannot be explained by the slightly lower coverage of the LNCaP exome. Most likely, these additional C4-2B mutations have arisen during tumor progression and bone metastasis.

### *4.4.2. Detecting point mutations in transcriptome sequencing*

Transcriptome sequencing was performed initially to determine differential gene expression. RNA was isolated from LNCaP and C4-2B cells that had been treated for 18 hours with the synthetic androgen methyltrienolone. We obtained 157 and 131 million 100 base pair, paired-end reads for LNCaP and C4-2B cells. In these reads, the percentage of ribosomal, intronic and intergenic bases was very low (1.6% in total), resulting in a high coverage of mRNA bases (Supplementary table 4.1). As a measure for the quality of the transcriptome data, the variation in coverage along each transcript is shown in Supplementary figure 4.2. This shows no 5' or 3' bias although there is a somewhat lower coverage near the ends of the transcripts.

The workflow for the detection and subsequent filtering of point mutations was similar to that used for the exome sequencing described higher. We found 1505 and 1882 mutations in LNCaP and C4-2B cells respectively, of which 1054 were detected in both cell lines (Supplementary table 4.5); 451 were specific for LNCaP and 828 for C4-2B.

### *4.4.3. Comparing exome with transcriptome sequencing data*

Comparing the read counts of variant alleles from genome and transcriptome sequencing data of all detected point mutations can be used as a measure of the sequencing quality. The majority of the mutations have a similar allele frequency in both DNA and RNA sequencing (Figure 4.1.A-B). Even the few homozygous mutations with allele frequency close to 1 in the exome data, have a similar allele frequency in the RNA sequencing data.

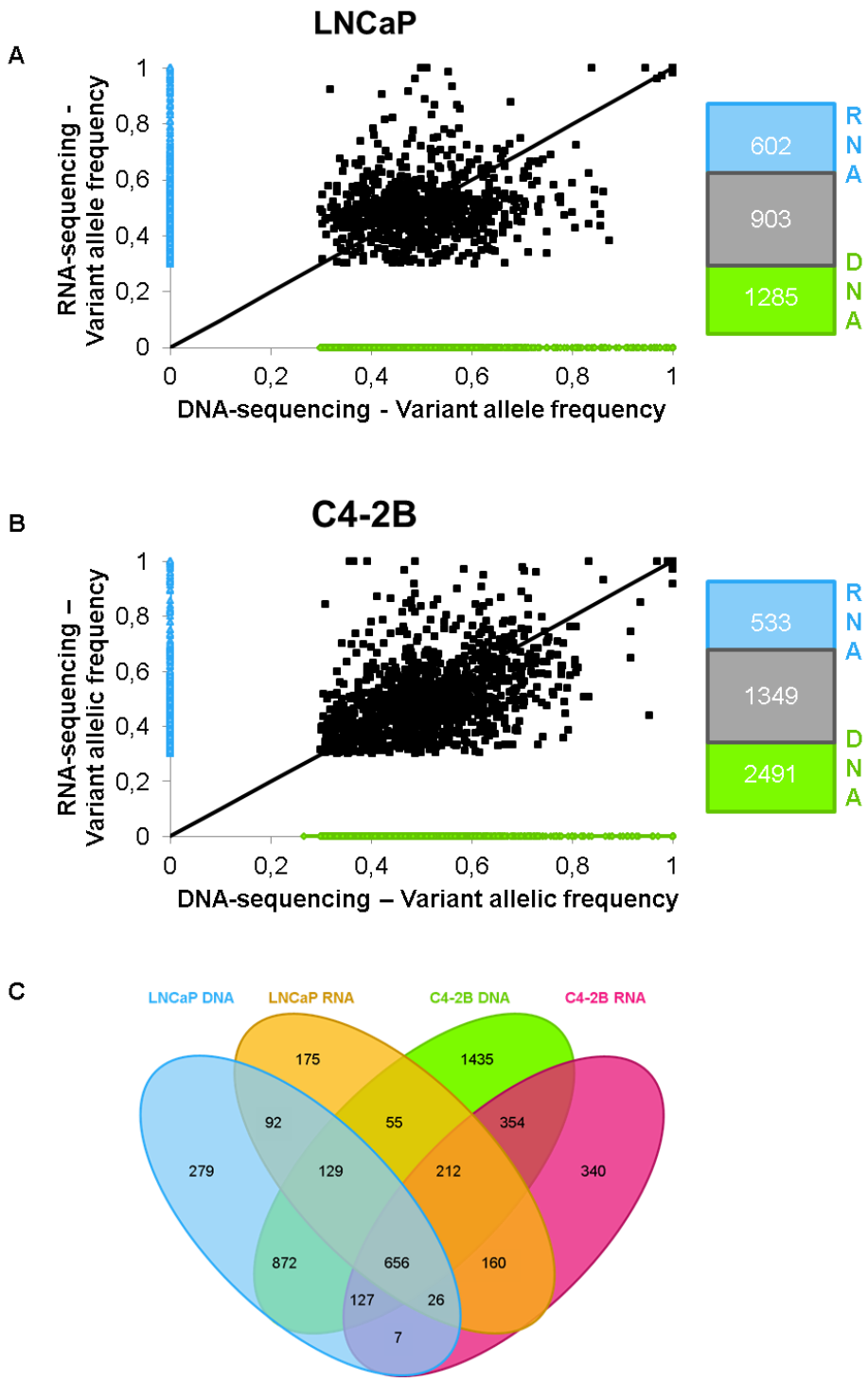



**Figure 4.1.** Comparison of variant allelic frequencies measured by whole exome and transcriptome sequencing. **A-B.** Comparison of the variant allelic frequency of mutations detected using whole exome and transcriptome sequencing. Black dots are mutations that have been found by both exome and transcriptome sequencing; red dots were only detected by exome sequencing and green dots only by RNA sequencing. For variant calling, a cut-off of 30% variant allelic frequency was applied. Next to the graph a figure shows the number of mutations from exome sequencing, transcriptome sequencing and the number found by both methods. Graphs are shown for LNCaP (A) and C4-2B cells (B) respectively. **C.** Overlap of all mutations observed by exome and transcriptome sequencing in LNCaP and C4-2B cells.

The combination of both the exome and transcriptome sequencing resulted in a total of 2244 mutations common to both cell lines (Figure 4.1.C). Moreover, the number of LNCaP-specific mutations (546) is much lower than that of C4-2B-specific changes (2129), again indicating that mutations have accumulated during the progression to C4-2B. RNA-sequencing confirmed only 41 and 35% of the exonic variants identified by whole exome sequencing of LNCaP and C4-2B. This number rose to 52% when we only took the expressed genes into account (FPKM > 1). Conversely, 60 and 71% of the LNCaP and C4-2B variants identified by transcriptome sequencing respectively were confirmed by exome sequencing.

### *4.4.4. Nucleotide substitutions*

The different types of transitions and transversions in the exomes and transcriptomes of LNCaP and C4-2B cell lines might give insight in the mutational processes that took place during the development of these cells. We observed that the predominant mutations (40-42%) in both cell lines were G-to-A and C-to-T transitions (Figure 4.2).

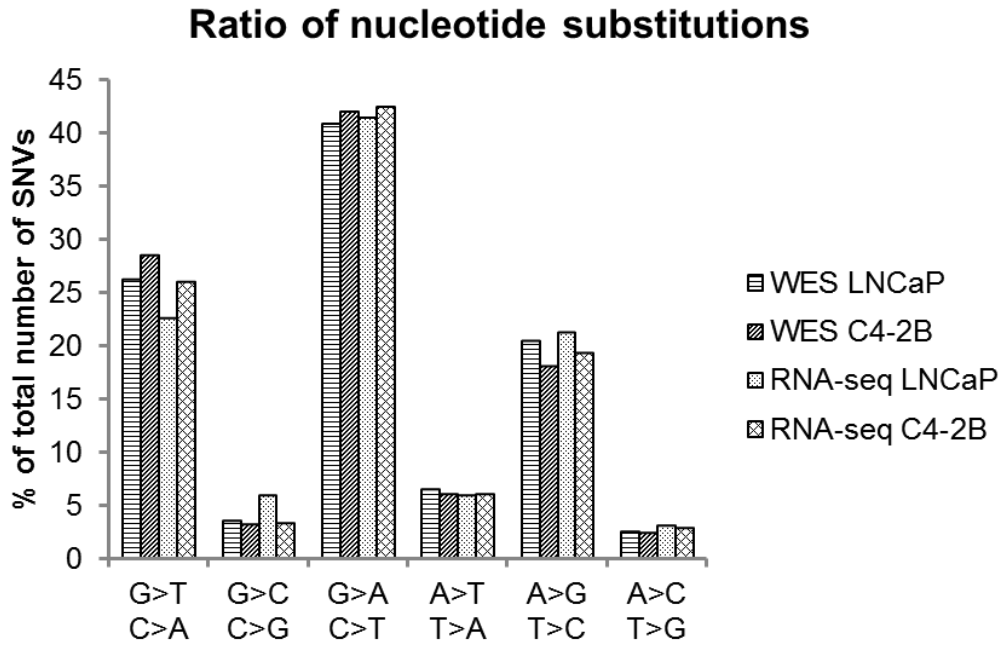



**Figure 4.2.** Mutation spectrum of single nucleotide substitutions. Percentages of mutations in each of the six possible mutation classes are represented for the exome sequencing and transcriptome sequencing data of both LNCaP and C4-2B cells.

The most prevalent type of RNA editing in higher eukaryotes is the conversion of adenosine to inosine. As inosine is read as a guanine after sequencing, this editing type manifests itself in RNA-sequencing as an A-to-G substitution (Dominissini *et al.* 2011). However, in our data sets, the number of A-to-G transitions in the exome and the transcriptome sequencing data is comparable arguing against an important role of RNA editing (Figure 4.2).

### *4.4.5. Validation of point mutations*

In total, 80 mutations in the exome data from LNCaP (47) and C4-2B (33) were validated by manual Sanger re-sequencing (Figure 4.3). The genes that were chosen for validation were ranked high in a functional prioritization of all mutated genes in the C4-2B cell line (calculated as in (Spans *et al.* 2012)). Nine of these mutations (in PIK3R1, TP53BP1, PRKCQ, CHEK2, RIPK2 and KLK3) were detected by DNA and RNA sequencing in both cell lines, and these were confirmed with Sanger sequencing on genomic and complementary DNA of LNCaP and C4-2B. When we tested seven of the C4-2B exome mutations (PIAS1 P216S and K380M, MKNK2 L229M and T244N, STAT5A I85N and MYO18A A1571T and Q646*), they were not detected by LNCaP exome sequencing, but their presence in the LNCaP genome was evident in the RNA sequencing data and also confirmed by Sanger sequencing on genomic and complementary DNA.

We also detected and confirmed C4-2B specific mutations in CASP9, FLNB, POLR2A and STAT5A in genomic DNA and cDNA of C4-2B cells, but not in LNCaP cells. Finally, mutations in genes that are not expressed in LNCaP or C4-2B (KIT and GRAP) could only be confirmed on genomic DNA.

In conclusion, the GATK UnifiedGenotyper for variant calling which we combined with our extensive filtering generated few false positives. Similar results were shown recently by Liu *et al.* by comparing GATK with SAMtools, Atlas 2 and glftools (Liu *et al.* 2013). Moreover, it should be noted that our validations indicated that the exome analyses did not uncover all mutations, but the variations that were discovered most likely are genuine mutations.

| | | LNCaP | | | | C4-2B | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Illumina | | Sanger | | Illumina | | Sanger | |
| | | DNA | RNA | DNA | RNA | DNA | RNA | DNA | RNA |
| *PIK3R1* | R369* | + | + | + | + | + | + | + | + |
| *TP53BP1* | Q111* | + | + | + | + | + | + | + | + |
| *TP53BP1* | R639Q | + | + | + | + | + | + | + | + |
| *PRKCQ* | E313* | + | + | + | + | + | + | + | + |
| *CHEK2* | T387N | + | + | + | + | + | + | + | + |
| *RIPK2* | H144Q | + | + | + | + | + | + | + | + |
| *KLK3* | P97T | + | + | + | + | + | + | + | + |
| *KLK3* | V136M | + | + | + | + | + | + | + | + |
| *KLK3* | L218P | + | + | + | + | + | + | + | + |
| *PIAS1* | P216S | / | + | + | + | + | + | + | + |
| *PIAS1* | K380M | / | + | + | + | + | + | + | + |
| *MKNK2* | L229M | - | + | + | + | + | + | + | + |
| *MKNK2* | T244N | - | + | + | + | + | + | + | + |
| *STAT5A* | I85N | - | + | + | + | + | + | + | + |
| *MYO18A* | A1571T | - | + | + | + | + | + | + | + |
| *MYO18A* | Q646* | - | - | + | + | + | + | + | + |
| *CASP9* | G161S | - | - | - | - | + | + | + | + |
| *FLNB* | S856N | - | - | - | - | + | + | + | + |
| *POLR2A* | T1184M | - | - | - | - | + | + | + | + |
| *STAT5A* | L656N | - | - | - | - | + | - | + | + |
| *PDCD6IP* | Y153* | - | - | - | - | + | - | + | - |
| *CUX1* | E1236* | - | - | - | - | + | - | - | - |
| *KIT* | L813I | - | / | + | 0 | + | / | + | 0 |
| *GRAP* | L128F | + | / | + | 0 | + | / | + | 0 |
| *SMARCA4* | L302P | - | - | - | - | + | + | - | - |
| *SPTBN1* | T802M | - | - | - | 0 | + | + | - | 0 |
| *SPTBN1* | W1044* | - | - | + | 0 | + | + | + | 0 |
| *MAPK4* | E106* | - | / | - | - | + | / | - | - |

**Figure 4.3.** Validation of point mutations in the LNCaP and C4-2B cell lines. Validation was performed using PCR on both genomic DNA and copy-DNA, followed by conventional Sanger sequencing. The first and second column represents the name of the gene and the amino acid substitution respectively. The third, fourth, seventh and eighth column represent next-generation sequencing results for whole exome and transcriptome sequencing, which are then validated with Sanger sequencing in the fifth, sixth, ninth and tenth column. A + denotes that the mutation was detected, a – denotes that the mutation was not detected, while 0 means that no PCR-product could be amplified and / that this position was not covered with next-generation sequencing.

### *4.4.6. Differential gene expression between LNCaP and C4-2B cells*

We next wanted to search for differentially expressed genes between the two cell lines, since these might provide clues to the mechanisms behind the evolution of LNCaP cells into C4-2B cells. Differential expression was called by the Tuxedo algorithm based on RNA-seq data of triplicates for each cell line, with additional filtering of log2-fold change > 2 and q-value < 0.001. All replicates were very similar, as can be seen in Supplementary figure 4.3. Moreover, the squared coefficient of variation, which is a normalized measure of cross-replicate variability, is below 0.05 for expressed genes.

Our analysis resulted in the identification of 703 differentially expressed genes (Supplementary table 4.6), of which 457 genes are higher expressed in C4-2B and 246 are higher expressed in LNCaP cells (Supplementary figure 4.2). An overview of the location of differentially expressed genes across the genome can be found in Figure 4.4, together with the frequency of point mutations detected in both cell lines, in C4-2B cells only, or in LNCaP cells only. Quantitative RT-PCR on five differentially expressed genes confirmed the RNA-seq data (data not shown).

Fu *et al.* already described some differentially expressed genes between LNCaP and C4-2B, but none of the genes they detected are differentially expressed in our data (Fu *et al.* 2002). We propose that culture conditions and differences between detection platforms most likely explain this discrepancy. On the other hand, there is considerable overlap of our datasets with those of other studies that compared LNCaP and C4-2 transcriptomes (Bisoffi *et al.* 2004, Liu *et al.* 2004, Oudes *et al.* 2005, Trojan *et al.* 2005, Xie *et al.* 2011).

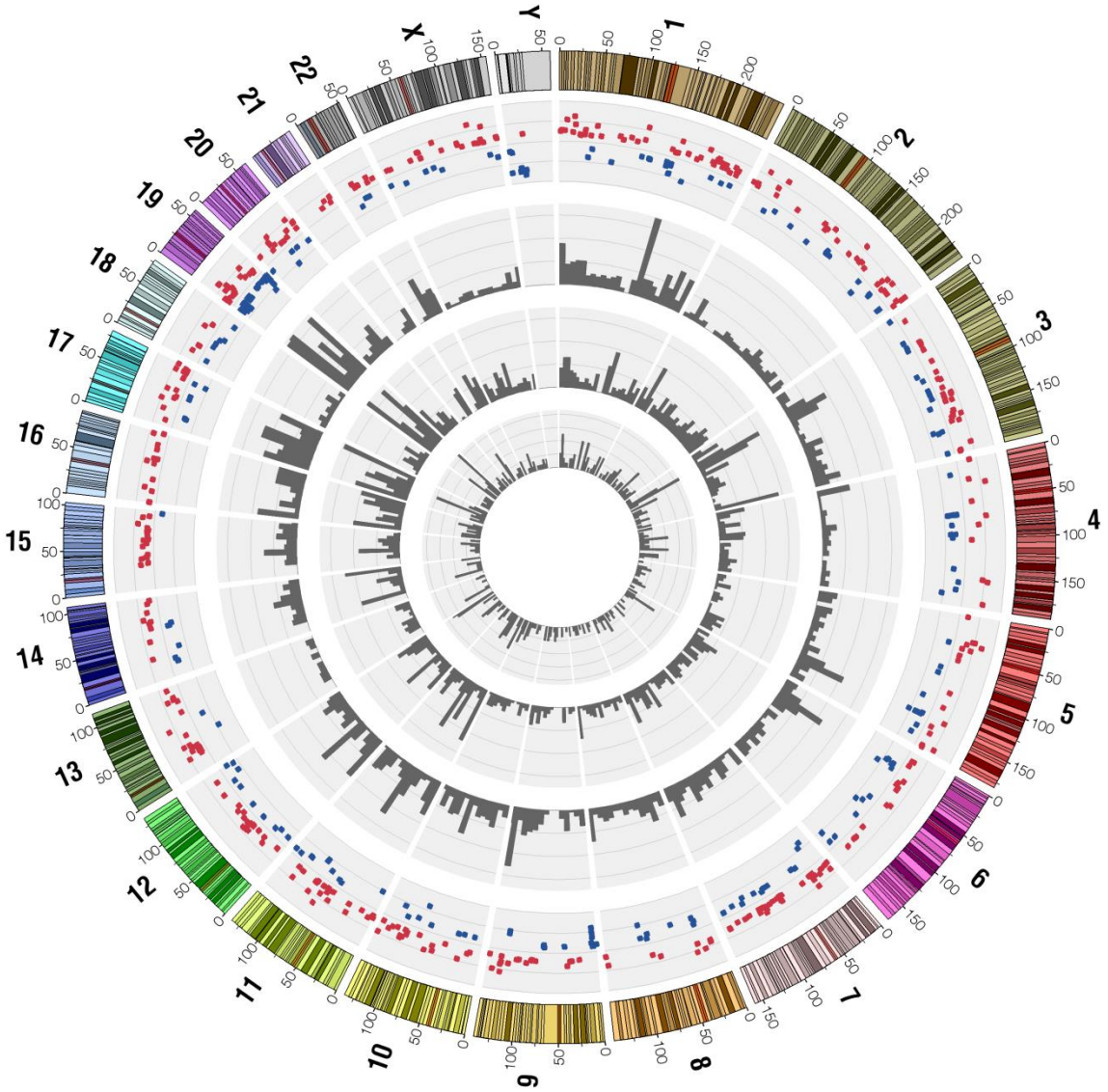


**Figure 4.4.** Representation of the catalog of differentially expressed genes and mutations in LNCaP and C4-2B cells. Chromosome ideograms are shown around the outer ring and oriented pter-qter in a clockwise direction with centromeres indicated in red. The outer ring represents differentially expressed genes: 457 genes with higher expression in C4-2B (red dots) and 246 genes with higher expression in LNCaP (blue dots). Other tracks contain (from outside to inside): 2244 mutations in common between LNCaP and C4-2B cells, 2129 mutations specific for C4-2B cells and 546 mutations specific for LNCaP cells shown by density per 10 megabases.

### *4.4.7. Pathway analysis of genomic and transcriptomic data sets*

LNCaP and C4-2B cells continue to be used in basic and preclinical research. We propose our databases of mutations and differentially expressed genes as important sources of inspiration for

further research projects. In addition, these databases can now be checked for specific mutations before one starts using these cells to study any specific PCa-related pathway.

This paragraph gives an example of a hypothesis based on *in silico* analysis of our data. Pathway-Express analysis of the C4-2B specific mutations combined with the 703 genes differentially expressed between LNCaP and C4-2B cells indicated that the most significant changes were found in the ECM-receptor interaction pathway and in focal adhesion. Both pathways converge in the upregulated expression of the myosin light chain kinase (MLCK) gene (Figure 4.5).

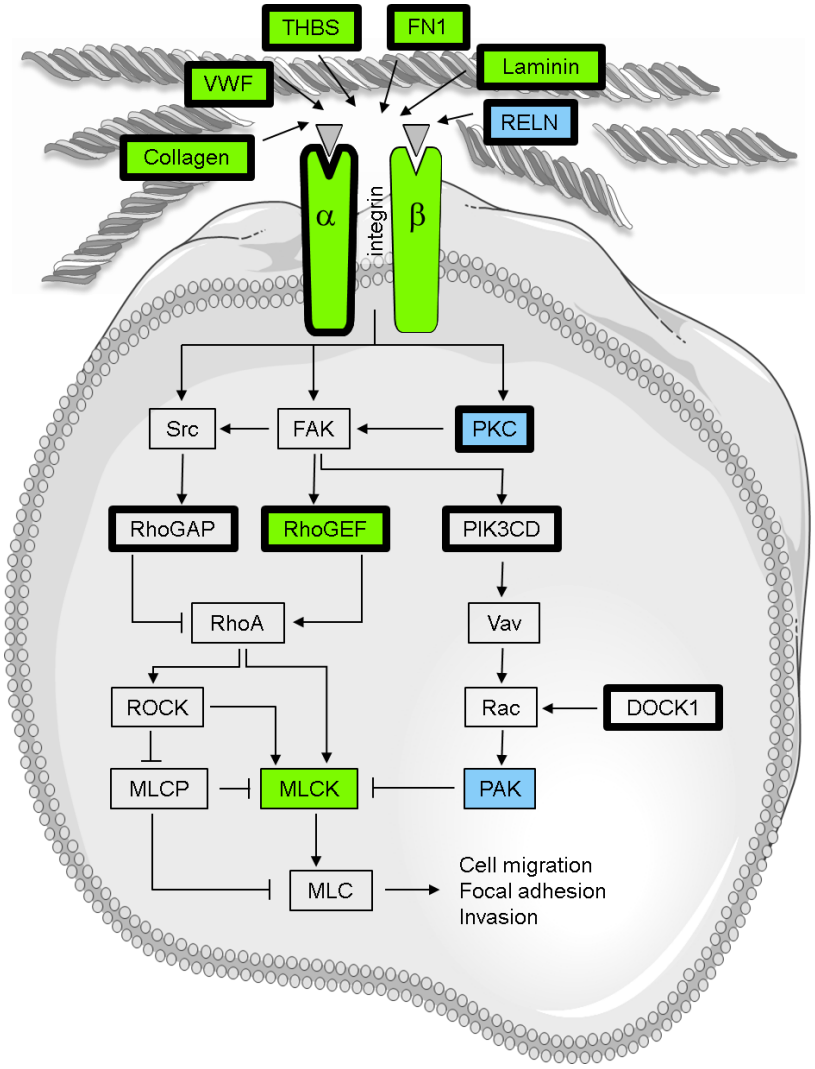



**Figure 4.5.** Alterations in the pathway converging on overexpression of Myosin Light Chain Kinase in C4-2B cells. Alterations are defined as having an increased (green) or decreased (blue) expression in C4-2B compared to LNCaP or by somatic mutations in C4-2B cells only (bold black lines). Overexpression of myosin light chain kinase in C4-2B cells might distinguish them from LNCaP cells in cell migration and focal adhesion characteristics.

## 4.5. DISCUSSION

### *4.5.1. A high mutation rate in LNCaP and C4-2B cells*

C4-2B cells are derived from a bone metastasis in nude mice inoculated with cells originating from the LNCaP-derived, castration-resistant xenografts called C4-2. They are considered a useful preclinical model for metastatic, castration-resistant and androgen receptor positive PCa. Here, we provide for the first time comparative maps of the point mutations detected in the LNCaP and C4-2B cells. In addition, although transcriptome analyses of LNCaP and C4-2 have been reported, to our knowledge, this is the first transcriptome analysis of C4-2B cells.

C4-2B cells as well as LNCaP cells have a surprisingly high number of point mutations: 4373 and 2790 mutations respectively. Like in primary PCa and castration-resistant PCa samples, the mutational spectrum is dominated by G-to-A and C-to-T transitions (Grasso *et al.* 2012, Lindberg *et al.* 2012, Weischenfeldt *et al.* 2013). It is known that mismatch repair defects cause transition mutations, particularly G-to-A and C-to-T substitutions (Lang *et al.* 2013). Hence, most mutations might be caused by the defective mismatch repair system in LNCaP cells, due to the homozygous deletion of the 3' end of the MSH2 gene (Leach *et al.* 2000). Chen *et al.* already described a correlating high instability of satellite DNA in LNCaP cells (Chen *et al.* 2001).

The number of point mutations in our cell lines is much higher than the average 16-33 mutations detected in whole exomes of PCa samples (Baca *et al.* 2013, Barbieri *et al.* 2012, Berger *et al.* 2011b, Weischenfeldt *et al.* 2013). These cell lines are therefore atypical, but might be considered a model for cases of PCa in which mismatch repair is defective as described for instance by Barbieri *et al.,* where a single PCa tumor harbored a frameshift mutation of the MSH6 gene among 996 other mutations (Barbieri *et al.* 2012). Obviously, such higher mutation rates would explain the even higher number of mutations we found in C4-2B compared to LNCaP. Unfortunately, this will also obscure the driver mutations that may have conferred a survival advantage during the metastatic process.

### *4.5.2. Link between mutation rates and expression*

For both the LNCaP and C4-2B cell line, we see that highly expressed genes more frequently contain point mutations than non-transcribed genes ($p < 0.0001$, Chi Square test, for the highest versus lowest expressed tertile). This contradicts the general link between heterochromatin organization and higher regional mutation rates in human cancer cells (Schuster-Bockler *et al.* 2012). Possibly, in these cell lines, the open chromatin and linked transcription induces more mismatches which normally are efficiently corrected, but not in case of a deficient mismatch repair.

### *4.5.3. Comparison of LNCaP and C4-2B mutations*

We detected 1784 shared mutations in the exomes of LNCaP and C4-2B, and 2056 C4-2B-specific changes, which makes sense since the C4-2B cells are derived from the LNCaP cells. However, we also detected 404 LNCaP-specific changes, many of which were confirmed by our transcriptome sequences. Obviously, the LNCaP cells we analyzed have deviated from the LNCaP cells that were used originally to develop the C4-2B cells (Thalmann *et al.* 1994). Indeed, we have shown earlier that even LNCaP cells from different labs are genetically different and while our cells were obtained from ATCC (passage 48), the C4-2B were most likely derived from a much earlier passage of LNCaP cells in 1994 (Spans *et al.* 2012, Thalmann *et al.* 1994).

### *4.5.4. Suggestion of a role of MLCK in the metastatic process*

Our data can clearly lead to the hypothesis on the metastatic process that took place during the conversion of LNCaP to C4-2B cells. This is exemplified by the convergence of a number of affected pathways to an upregulation of MLCK. Indeed, there are several published links between MLCK and the metastatic process. Discriminant analysis of microarrays identified the MLCK gene as the most informative gene for the PCa genesis process (Fujita *et al.* 2008), and inhibition of MLCK in rat PCa cells results in reduction of invasiveness, which was principally due to impaired cellular motility (Tohtong *et al.* 2003). Inhibiting MLCK in fibrosarcoma, pancreatic cancer and breast cancer cells also results in decreased adhesion, migration and invasion and increased apoptosis (Cui *et al.* 2010, Fazal *et al.* 2005, Kaneko *et al.* 2002, Niggli *et al.* 2006). Conversely, activating MLCK leads to an increase in invasion in breast cancer cells and an increased metastatic potential in non-small cell lung cancer (Khuon *et al.* 2010, Minamiya *et al.* 2005). The differential expression of the MLCK gene in the two cell lines investigated here might therefore correlate with the higher metastatic capacity of the C4-2B cells.

In conclusion, our data clearly show that there are major differences in the number and distribution of mutations and gene expression between LNCaP and C4-2B cells. Since these cell lines are universally used to study the progression from non-metastatic to metastatic PCa, these data are crucial for researchers to correctly interpret their results when using these cell lines. Moreover, our databases will be very helpful in developing new investigational ideas.

**4.6. REFERENCES**


Baca, S. C., Prandi, D., Lawrence, M. S*., et al.* (2013). "Punctuated evolution of prostate cancer genomes." Cell **153**(3): 666-677.

Barbieri, C. E., Baca, S. C., Lawrence, M. S*., et al.* (2012). "Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer." Nat Genet **44**(6): 685-689.

Berger, M. F., Lawrence, M. S., Demichelis, F*., et al.* (2011). "The genomic complexity of primary human prostate cancer." Nature **470**(7333): 214-220.

Bisoffi, M., Klima, I., Gresko, E*., et al.* (2004). "Expression profiles of androgen independent bone metastatic prostate cancer cells indicate up-regulation of the putative serine-threonine kinase GS3955." J Urol **172**(3): 1145-1150.

Chen, Y., Wang, J., Fraig, M. M*., et al.* (2001). "Defects of DNA mismatch repair in human prostate cancer." Cancer Res **61**(10): 4112-4121.

Cui, W. J., Liu, Y., Zhou, X. L*., et al.* (2010). "Myosin light chain kinase is responsible for high proliferative ability of breast cancer cells via anti-apoptosis involving p38 pathway." Acta Pharmacol Sin **31**(6): 725-732.

Dominissini, D., Moshitch-Moshkovitz, S., Amariglio, N*., et al.* (2011). "Adenosine-to-inosine RNA editing meets cancer." Carcinogenesis **32**(11): 1569-1577.

Draghici, S., Khatri, P., Tarca, A. L*., et al.* (2007). "A systems biology approach for pathway level analysis." Genome Res **17**(10): 1537-1545.

Fazal, F., Gu, L., Ihnatovych, I*., et al.* (2005). "Inhibiting myosin light chain kinase induces apoptosis in vitro and in vivo." Mol Cell Biol **25**(14): 6259-6266.

Ferlay, J., Steliarova-Foucher, E., Lortet-Tieulent, J*., et al.* (2013). "Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012." Eur J Cancer **49**(6): 1374-1403.

Fu, Z., Dozmorov, I. M. and Keller, E. T. (2002). "Osteoblasts produce soluble factors that induce a gene expression pattern in non-metastatic prostate cancer cells, similar to that found in bone metastatic prostate cancer cells." Prostate **51**(1): 10-20.

Fujita, A., Gomes, L. R., Sato, J. R*., et al.* (2008). "Multivariate gene expression analysis reveals functional connectivity changes between normal/tumoral prostates." BMC Syst Biol **2**: 106.

Grasso, C. S., Wu, Y. M., Robinson, D. R*., et al.* (2012). "The mutational landscape of lethal castration-resistant prostate cancer." Nature **487**(7406): 239-243.

Haffner, M. C., Mosbruger, T., Esopi, D. M*., et al.* (2013). "Tracking the clonal origin of lethal prostate cancer." J Clin Invest **123**(11): 4918-4922.

Horoszewicz, J. S., Leong, S. S., Chu, T. M*., et al.* (1980). "The LNCaP cell line--a new model for studies on human prostatic carcinoma." Prog Clin Biol Res **37**: 115-132.

Ianculescu, I., Wu, D. Y., Siegmund, K. D*., et al.* (2012). "Selective roles for cAMP response element-binding protein binding protein and p300 protein as coregulators for androgen-regulated gene expression in advanced prostate cancer cells." J Biol Chem **287**(6): 4000-4013.

Kaneko, K., Satoh, K., Masamune, A*., et al.* (2002). "Myosin light chain kinase inhibitors can block invasion and adhesion of human pancreatic cancer cell lines." Pancreas **24**(1): 34-41.

Khuon, S., Liang, L., Dettman, R. W*., et al.* (2010). "Myosin light chain kinase mediates transcellular intravasation of breast cancer cells through the underlying endothelial cells: a three-dimensional FRET study." J Cell Sci **123**(Pt 3): 431-440.

Krzywinski, M., Schein, J., Birol, I*., et al.* (2009). "Circos: an information aesthetic for comparative genomics." Genome Res **19**(9): 1639-1645.

Lang, G. I., Parsons, L. and Gammie, A. E. (2013). "Mutation rates, spectra, and genome-wide distribution of spontaneous mutations in mismatch repair deficient yeast." G3 (Bethesda) **3**(9): 1453-1465.

Leach, F. S., Velasco, A., Hsieh, J. T*., et al.* (2000). "The mismatch repair gene hMSH2 is mutated in the prostate cancer cell line LNCaP." J Urol **164**(5): 1830-1833.

Li, H. and Durbin, R. (2009). "Fast and accurate short read alignment with Burrows-Wheeler transform." Bioinformatics **25**(14): 1754-1760.

Lindberg, J., Mills, I. G., Klevebring, D.*, et al.* (2013). "The mitochondrial and autosomal mutation landscapes of prostate cancer." Eur Urol **63**(4): 702-708.

Liu, A. Y., Brubaker, K. D., Goo, Y. A.*, et al.* (2004). "Lineage relationship between LNCaP and LNCaP-derived prostate cancer cell lines." Prostate **60**(2): 98-108.

Liu, X., Han, S., Wang, Z.*, et al.* (2013). "Variant callers for next-generation sequencing data: a comparison study." PLoS One **8**(9): e75619.

Lu-Yao, G. L., Albertsen, P. C., Moore, D. F.*, et al.* (2009). "Outcomes of localized prostate cancer following conservative management." JAMA **302**(11): 1202-1209.

McKenna, A., Hanna, M., Banks, E.*, et al.* (2010). "The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data." Genome Res **20**(9): 1297-1303.

Minamiya, Y., Nakagawa, T., Saito, H.*, et al.* (2005). "Increased expression of myosin light chain kinase mRNA is related to metastasis in non-small cell lung cancer." Tumour Biol **26**(3): 153-157.

Niggli, V., Schmid, M. and Nievergelt, A. (2006). "Differential roles of Rho-kinase and myosin light chain kinase in regulating shape, adhesion, and migration of HT1080 fibrosarcoma cells." Biochem Biophys Res Commun **343**(2): 602-608.

Oudes, A. J., Roach, J. C., Walashek, L. S.*, et al.* (2005). "Application of Affymetrix array and Massively Parallel Signature Sequencing for identification of genes involved in prostate cancer progression." BMC Cancer **5**: 86.

Pan, Y., Kytola, S., Farnebo, F.*, et al.* (1999). "Characterization of chromosomal abnormalities in prostate cancer cell lines by spectral karyotyping." Cytogenet Cell Genet **87**(3-4): 225-232.

Rider, J. R., Sandin, F., Andren, O.*, et al.* (2013). "Long-term outcomes among noncuratively treated men according to prostate cancer risk category in a nationwide, population-based study." Eur Urol **63**(1): 88-96.

Sampson, N., Neuwirt, H., Puhr, M.*, et al.* (2013). "In vitro model systems to study androgen receptor signaling in prostate cancer." Endocr Relat Cancer **20**(2): R49-64.

Schuster-Bockler, B. and Lehner, B. (2012). "Chromatin organization is a major influence on regional mutation rates in human cancer cells." Nature **488**(7412): 504-507.

Siegel, R., DeSantis, C., Virgo, K.*, et al.* (2012). "Cancer treatment and survivorship statistics, 2012." CA Cancer J Clin **62**(4): 220-241.

Spans, L., Atak, Z. K., Van Nieuwerburgh, F.*, et al.* (2012). "Variations in the exome of the LNCaP prostate cancer cell line." Prostate **72**(12): 1317-1327.

Spans, L., Clinckemalie, L., Helsen, C.*, et al.* (2013). "The genomic landscape of prostate cancer." Int J Mol Sci **14**(6): 10822-10851.

Thalmann, G. N., Anezinis, P. E., Chang, S. M.*, et al.* (1994). "Androgen-independent cancer progression and bone metastasis in the LNCaP model of human prostate cancer." Cancer Res **54**(10): 2577-2581.

Tohtong, R., Phattarasakul, K., Jiraviriyakul, A.*, et al.* (2003). "Dependence of metastatic cancer cell invasion on MLCK-catalyzed phosphorylation of myosin regulatory light chain." Prostate Cancer Prostatic Dis **6**(3): 212-216.

Trapnell, C., Roberts, A., Goff, L.*, et al.* (2012). "Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks." Nat Protoc **7**(3): 562-578.

Trojan, L., Schaaf, A., Steidler, A.*, et al.* (2005). "Identification of metastasis-associated genes in prostate cancer by genetic profiling of human prostate cancer cell lines." Anticancer Res **25**(1A): 183-191.

Weischenfeldt, J., Simon, R., Feuerbach, L.*, et al.* (2013). "Integrative genomic analyses reveal an androgen-driven somatic alteration landscape in early-onset prostate cancer." Cancer Cell **23**(2): 159-170.

Xie, B. X., Zhang, H., Wang, J.*, et al.* (2011). "Analysis of differentially expressed genes in LNCaP prostate cancer progression model." J Androl **32**(2): 170-182.

## 4.7. SUPPLEMENTARY INFORMATION

**Supplementary table 4.1.** Sequencing characteristics

**Supplementary table 4.2.** Filters used to identify point mutations in the exomes of LNCaP and C4-2B cells

**Supplementary table 4.3.** List of 2188 non-synonymous point mutations detected in LNCaP cells

The table is not included in this thesis due to space constraints and can be obtained through the published article.

**Supplementary table 4.4.** List of 3840 non-synonymous point mutations detected in C4-2B cells

The table is not included in this thesis due to space constraints and can be obtained through the published article.

**Supplementary table 4.5.** Filters used to identify point mutations in the transcriptomes of LNCaP and C4-2B cells

**Supplementary table 4.6.** List of 703 genes differentially expressed between LNCaP and C4-2B cells

The table is not included in this thesis due to space constraints and can be obtained through the published article.

**Supplementary figure 4.1.** FastQC quality control results of the per base qualities

**Supplementary figure 4.2.** Normalized coverage by position

**Supplementary figure 4.3.** Heatmap of 703 differentially expressed genes

**Supplementary table 4.1.** Sequencing characteristics

| | ***Whole exome sequencing*** | | ***RNA sequencing*** | |
|---|---|---|---|---|
| | ***LNCaP*** | ***C4-2B*** | ***LNCaP*** | ***C4-2B*** |
| Total reads | 49790757 | 80299454 | 157583847 | 131925448 |
| Aligned HQ reads | 34791304 | 73388716 | 155837048 | 130536483 |
| Aligned bases | 3454982953 | 7455182568 | 15915376614 | 13323971443 |
| HQ aligned Q20 bases | 3085390366 | 6710948381 | 15016055477 | 12559510784 |
| Percentage aligned reads that passed filter | 76,70% | 98,00% | 100% | 100% |
| Mean read length | 101 | 101 | 101 | 101 |
| | | | | |
| Zero coverage targets | 6,33% | 1,52% | | |
| Percentage target bases 10x | 79,04% | 93,18% | | |
| Percentage target bases 20x | 74,56% | 88,84% | | |
| | | | | |
| Percentage ribosomal bases | | | 0,0001% | 0,0001% |
| Percentage UTR bases | | | 32,4079% | 30,8155% |
| Percentage intronic bases | | | 0,7241% | 0,7388% |
| Percentage intergenic bases | | | 0,9193% | 0,9151% |
| Percentage mRNA bases | | | 98,3566% | 98,3459% |

**Supplementary table 4.2.** Filters used to identify point mutations in the exomes of LNCaP and C4-2B cells

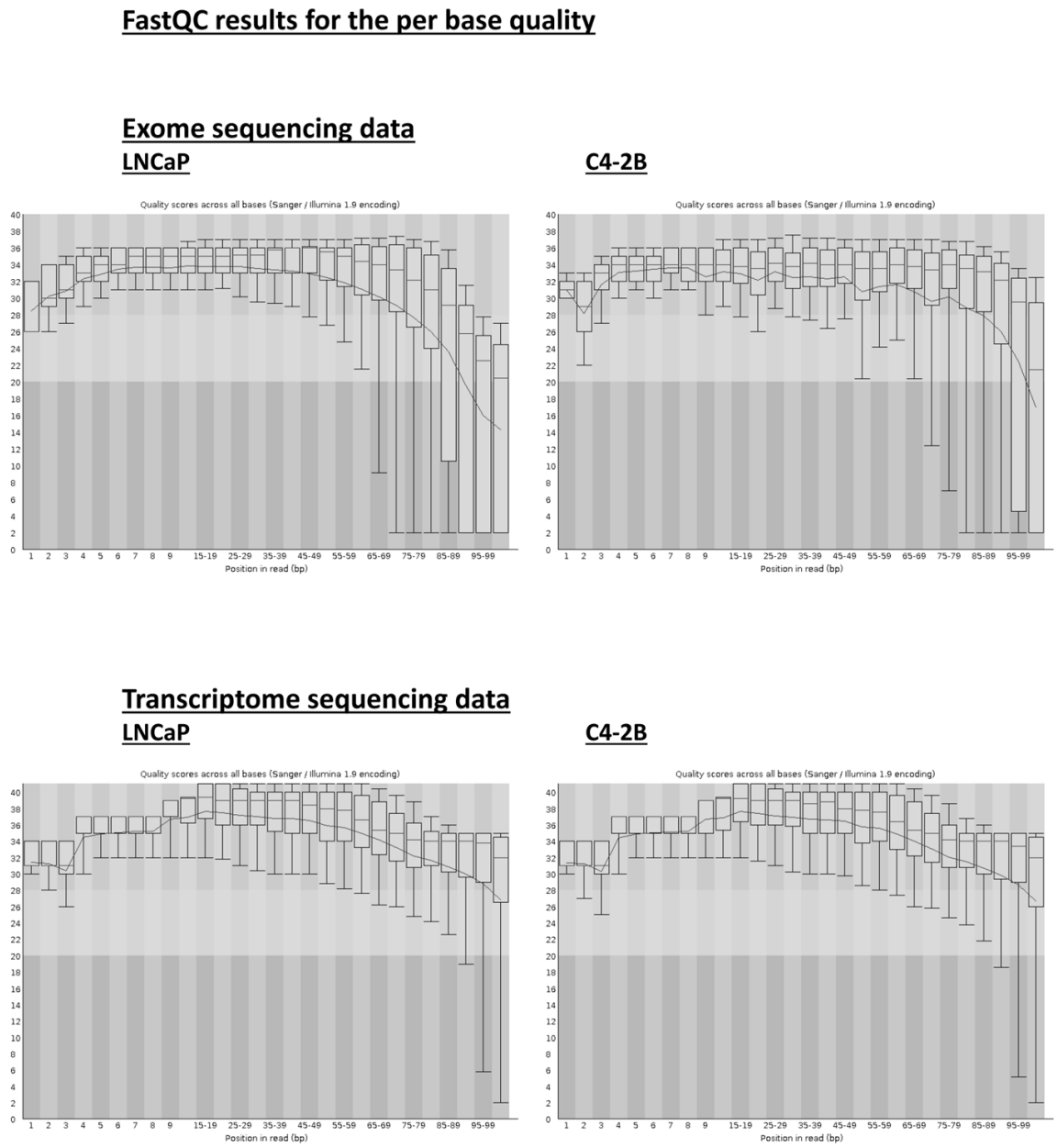
| ***Number of SNVs*** | ***LNCaP*** | ***C4-2B*** |
|---|---|---|
| Predicted | 514 295 | 380 691 |
| Not 'lowqual' according to GATK | 143 004 | 157 708 |
| Not present in dbSNP130 | 23 896 | 38 648 |
| With coverage ≥ 12x | 10 165 | 29 450 |
| With mutation allelic frequency ≥ 30% | 7 898 | 18 718 |
| Nonsense or missense | 2 302 | 4 562 |
| Without strand bias | 2 188 | 3 840 |
| In common | 1 784 | |

Mutation allelic frequency denotes the number of reads containing the mutation divided by the total number of reads.

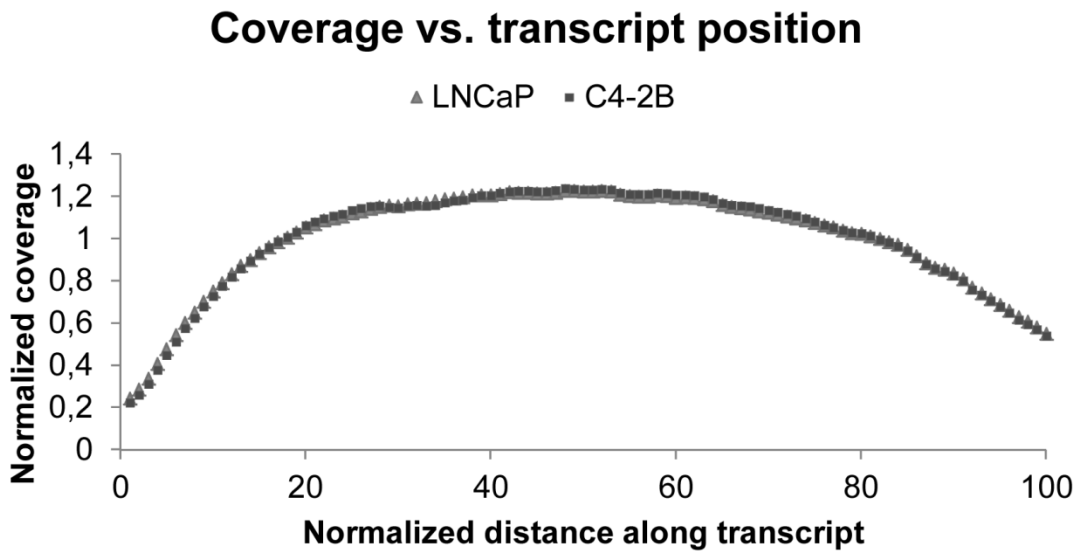
**Supplementary table 4.5.** Filters used to identify point mutations in the transcriptomes of LNCaP and C4-2B cells

| ***Number of SNVs*** | ***LNCaP*** | ***C4-2B*** |
|---|---|---|
| Predicted | 27 748 | 32 594 |
| Not 'lowqual' according to GATK | 15 250 | 18 655 |
| With coverage ≥ 12x | 12 532 | 15 807 |
| With mutation allelic frequency ≥ 30% | 6 872 | 7 669 |
| Not present in dbSNP130 | 6 802 | 7 606 |
| Nonsense or missense | 1 551 | 1 932 |
| Without strand bias | 1 606 | 1 881 |
| In common | 1 054 | |

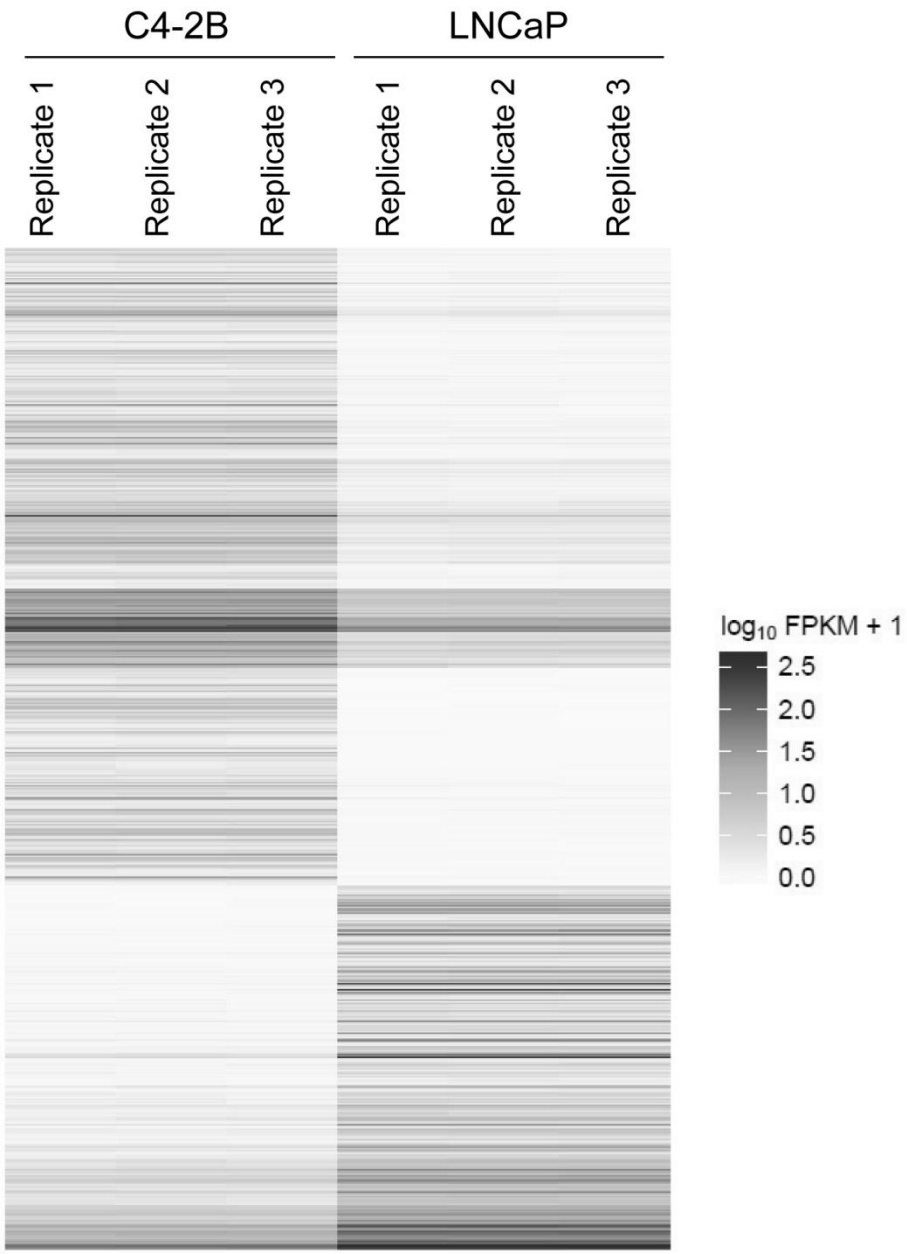
Mutation allelic frequency denotes the number of reads containing the mutation divided by the total number of reads.




**Supplementary figure 4.1.** FastQC quality control results of the per base qualities. Output results of the FastQC quality control software (version 0.10.1) are shown here for exome and transcriptome sequencing of LNCaP and C4-2B cells.




**Supplementary figure 4.2.** Normalized coverage by position. The average relative coverage is shown at each relative position along the transcript's length. LNCaP is depicted in green, while C4-2B is depicted in red. The x-axis represents the gene length normalized to 100%, where 0 is the 5' end of each transcript and 100 is the 3' end.




**Supplementary figure 4.3.** Heatmap of 703 differentially expressed genes. The heatmap shows the three replicates of each cell line, which are very similar. All differentially expressed genes were detected using the Tuxedo algorithm, with $q < 0.001$ and log2-fold change > 2 as cut-offs. It is clear that the majority of genes is upregulated in C4-2B compared to LNCaP, while a smaller group of genes is downregulated in C4-2B.

## CHAPTER 5:

## GENOMIC ANALYSIS OF HIGH-RISK PROSTATE CANCER REVEALS LOSS-OF-FUNCTION MUTATIONS IN RFC1 AND TET1

The work presented in this chapter has been conducted by the PhD candidate, unless otherwise stated in the figure legends.

### 5.1. ABSTRACT

Prostate cancer is the second most frequently diagnosed cancer in males worldwide. A wide range of genomic alterations, including point mutations, copy number changes and rearrangements, can lead to the development of cancer. Here, we report on the exome sequencing and copy number aberrations of 27 high-risk primary prostate tumors and their normal tissue pairs. These prostate cancer samples were collected within the PEARL consortium (ProstatE cAncer Research team Leuven). In addition to the amplifications and deletions that were described before, we identified a novel amplification on 7p22.3. Exome sequencing revealed one hypermutated sample containing 451 mutations, compared to an average of 19 mutations in the other samples, indicating that DNA repair is compromised. This hypermutated tumor harbored a mutation in the DNA-repair gene Replication Factor C (RFC1). The mutation is predicted to affect the interaction with PCNA and hence the recruitment of DNA polymerases to PCNA. In a second tumor, we detected a novel point mutation in the TET1 gene (Ten-Eleven Translocation 1). This methylcytosine dioxygenase converts 5-methylcytosine to 5-hydroxymethylcytosine, which leads to demethylation of cytosine and might lead to gene expression changes or altered chromatin organization. MeDIP-Seq and hMeDIP-Seq experiments performed on the tumor sample containing the A1908S TET1 mutation demonstrated an overall hypo-hydroxymethylation and hypermethylation at specific genomic loci, when compared to two other tumor samples without mutation in TET1. This effect was corroborated by the *in vitro* effect of the mutation on the dioxygenase activity as assessed by dot blot assays. We further showed that there is an overlap between the groups of androgen-regulated genes and the TET1-regulated genes. Moreover, TET1 is a coactivator of the androgen receptor. From the above data, we conclude that the A1908S TET1 mutation as detected in a high-risk primary prostate cancer leads to partial loss of TET1 tumor suppressor activity.

## 5.2. INTRODUCTION

Prostate cancer (PCa) is the most frequently diagnosed cancer amongst males in Belgium and in Europe (Ferlay *et al.* 2013, Hsu *et al.* 2012). In Belgium, 9036 new cases were diagnosed in 2011 (www.kankerregister.org). There is a large, uncharacterized clinical heterogeneity between patients diagnosed with PCa, with some patients dying of metastatic disease within two or three years after diagnosis, while other patients can live for 20 years with organ-confined disease. This clinical heterogeneity is most likely a reflection of the underlying genetic diversity. To obtain a better understanding of these genomic changes, we need large scale PCa genome characterization projects. Due to the relatively small tumor size and the admixture of stroma and non-tumoral tissue, such projects are not as easy for PCa as they are for other cancers. Nevertheless, the contribution of somatic base pair mutations to the oncogenic process has recently been studied in large cohorts, for both primary and metastatic PCa (Barbieri *et al.* 2012, Grasso *et al.* 2012). In addition to well-known dysregulated genes involved in chromatin modification, cell-cycle regulation and androgen signaling, they identified novel recurrently mutated genes such as SPOP in primary PCa and FOXA1 in castration-resistant PCa (CRPC). Despite the contributions of these large studies, new substitutions are still discovered with almost every new case that is analyzed.

Up to 25% of newly diagnosed patients have high-risk PCa, meaning that they have a high risk of recurrence of the disease and a high risk of dying from PCa (Meng *et al.* 2005). High-risk patients are defined as having at least one of the following risk factors: PSA ≥ 20 ng/ml, biopsy Gleason score ≥ 8, clinical T-stage ≥ T2c (D'Amico *et al.* 1998). However, these patients have heterogeneous outcomes: not all patients have invariably poor prognoses and the challenge is to better define lethal PCa (Briganti *et al.* 2012). Therefore, we need to understand the underlying genetic changes that can diversify patients according to the aggressiveness of the disease. The present study focused only on high-risk PCa patients, in the hope it would eventually lead to novel classifications of PCa patients by adding genetic markers to the clinical diagnostics and prognostics.

DNA methylation plays an important role in development, aging and disease. Indeed, the methylation of cytosine to methylcytosine (5mC) is so crucial that is has been called the fifth base in the genome. Despite the discovery of DNA methylation more than 60 years ago, and its known role in most processes involving the genome, the mechanisms controlling 5mC dynamics are only starting to being unraveled (Hotchkiss 1948). In 2009, TET1 (ten-eleven translocation 1) was identified as a dioxygenase that catalyzes the conversion of 5mC to 5-hydroxymethylcytosine (5hmC), also known as the sixth base (Tahiliani *et al.* 2009). The latter can be further oxidized to 5-formylcytosine (5fC) and

5-carboxylcytosine (5caC) (He *et al.* 2011, Ito *et al.* 2011). This led to the discovery that the activity of TET1, as well as the related TET2 and TET3 proteins, participate in active or passive DNA demethylation (Figure 5.1) (Guo *et al.* 2011). Passive demethylation of 5mC oxidation products can occur through cell division, while active demethylation consists of several steps including the oxidation of 5hmC to 5fC or 5caC or the deamination of 5hmC to 5hmU (5-hydroxymethyluracil) followed by the removal of these modified bases by base excision repair (BER).

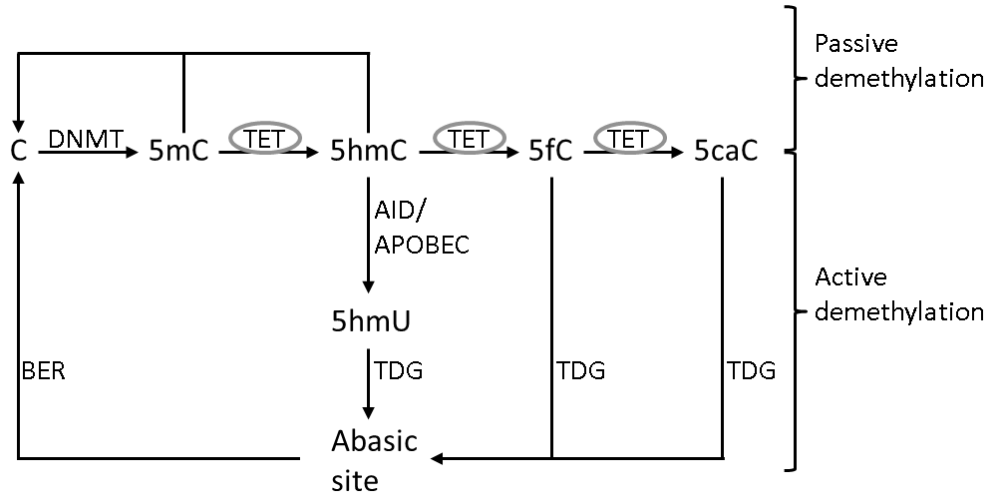



**Figure 5.1.** Involvement of TET family proteins in the regulation of DNA methylation and demethylation. Genomic 5mC can be removed passively during replication, but active demethylation has also been proposed. TET proteins convert 5mC to 5hmC, which can then be converted to 5hmU by AID/APOBEC enzymes. Consecutive action of TDG and BER result in an unmethylated cytosine. Abbreviations: DNMT, DNA methyl transferase; TET, ten-eleven translocation; AID, activation-induced deaminase; APOBEC, apolipoprotein B mRNA-editing enzyme complex; TDG, thymine DNA glycosylase; BER, base excision repair.

Patterns of DNA methylation are profoundly altered in human cancers. In particular, aberrant promoter hypermethylation leading to inappropriate transcriptional silencing of tumor suppressor genes and a global hypomethylation in intergenic regions are characteristic for cancer cells (Baylin *et al.* 2011). Loss of 5hmC has also been described in various solid tumors, including breast, colon, lung, kidney, and prostate cancer (Haffner *et al.* 2011, Jin *et al.* 2011, Kudo *et al.* 2012). Moreover, substantial downregulation of all three TET genes, but most significantly of TET1, has been described in colorectal, liver, breast and prostate cancers (Kudo *et al.* 2012, Yang *et al.* 2013). In particular, 33% of 153 PCa patients showed a reduced TET1 expression in cancer cells as compared to non-neoplastic glands (Hsu *et al.* 2012). Finally, TET1 has been shown to be an essential tumor suppressor in PCa and breast cancer through the downregulated methylation of critical genes such as TIMP2 and TIMP3 (Hsu *et al.* 2012). Here we report a first mutation in the TET1 gene for high-risk PCa.

## 5.3. MATERIALS AND METHODS

### *5.3.1. Materials*

Oligonucleotides were purchased from Integrated DNA Technologies. PCR reactions were performed on a Progene thermocycler using Expand High Fidelity PCR System (Roche Applied Science). All plasmid constructs were purified using the Qiagen Plasmid Midi kit (Qiagen) and analyzed by Sanger sequencing analysis at LGC Genomics. Antibodies used for dot blot, DIP-Seq and immunohistochemistry are antibodies against 5-methylcytosine (Eurogentec) and 5-hydroxymethylcytosine (Active Motif).

### *5.3.2. Plasmid constructs*

The luciferase reporter plasmid and the androgen receptor (AR) expression vector have been described before (Denayer *et al.* 2010). The classical ARE-regulated luciferase reporter plasmid was based on a pGEM-4Z vector that contains four copies of the SLP-MUT ARE, a minimal E1B-TATA box, and the pGL4 luciferase gene (Promega), all surrounded by chicken β-globin insulators. The pCMV-β-gal expression plasmid was obtained from Stratagene. Constructs containing the catalytic domain of TET1 in a pEF1 vector were a kind gift from Kian Koh (Stem Cell Biology and Embryology Laboratory, Department of Development and Regeneration, KU Leuven). Full length TET1 in pcDNA6.2 plasmids were a kind gift from Diether Lambrechts (Vesalius Research Center, KU Leuven).

### *5.3.3. Prostate cancer sample acquisition*

High-risk primary prostate tumors were obtained under protocols approved by the KU Leuven Ethical Committee from consented patients undergoing radical prostatectomy. Matching germline DNA was derived either from histologically benign prostate tissue or peripheral blood cells. Specimens were collected at the University Hospitals of Leuven within the PEARL consortium (ProstatE cAncer Research team Leuven). After tissue homogenization and lysis, DNA was extracted and assessed for quality by Nanodrop and PicoGreen measurements.

### *5.3.4. Copy number analyses*

Genome-wide SNP genotyping was performed using Illumina CytoSNP arrays on an iSCAN (Illumina). Processing of DNA samples, hybridization, staining, scanning of the BeadChips, and primary data extraction were all performed according to the Illumina Infinium protocol at the Vesalius Research

Center (Leuven, Belgium). GenomeStudio software was used for primary assessment of data and quality control assessment. The ASCAT (Allele-Specific Copy number Analysis of Tumors) software (version 2.1) was used to accurately determine allele-specific copy number alterations from SNP array data in solid tumors, while estimating and correcting for both tumor aneuploidy and infiltration of non-aberrant cells (Van Loo *et al.* 2010). To identify genomic regions that were significantly amplified or deleted, the Genomic Identification of Significant Targets in Cancer (GISTIC, version 2.0.1) method was used, as previously described (Beroukhim *et al.* 2007).

### *5.3.5. Exome sequencing*

The exome capture was performed using the SeqCap EZ Exome version 2 kit (Roche Nimblegen), after which paired-end sequences of 101 nucleotides were generated with an Illumina HiSeq2000 or HiSeq2500 instrument (Genomics Core, UZ Leuven). Sequencing data were aligned to the hg19 human reference genome with BWA and processed by the Picard pipelines (Li *et al.* 2009a) (http://picard.sourceforge.net/). Alignment files were processed further with Genome Analysis Toolkit (GATK) before variant calling and included duplicate removal, local realignment around known indels and base quality recalibration (McKenna *et al.* 2010). Somatic point mutations were detected by comparison of tumor and paired normal exome sequences with MuTect and SomaticSniper (Cibulskis *et al.* 2013, Larson *et al.* 2012). Single nucleotide variants (SNVs) were annotated with the SeattleSeq tool (http://snp.gs.washington.edu/SeattleSeqAnnotation138/).

### *5.3.6. Validation of point mutations*

Validation of somatic missense and nonsense variants in the tumor and matched normal genomes was performed either using standard Sequenom MassARRAY genotyping experiments or Sanger sequencing. Sequenom validation was performed according to the manufacturer's conditions. Automated genotyping calls were generated using the MassARRAY RTTM software (Sequenom) and were validated by manual review of the raw mass spectra. The following approach was used: SNVs that failed to be successfully genotyped in the first round of validation were subsequently redesigned for a second attempt using a new set of Sequenom primers. When SNVs also failed to be successfully genotyped in this second round, they were considered as 'failed genotyping using Sequenom'. The Sequenom validation experiment was performed at the Vesalius Research Center (Leuven, Belgium). We also carried out Sanger sequencing to try to validate SNVs that could not be validated by Sequenom. We therefore designed primers specific to the region containing the SNV to be tested

using the NCBI Primer-Blast (http://ncbi.nlm.nih.gov/tools/primer-blast/). Standard PCR reactions were sequenced and sequence trace files were analyzed using Chromas Lite.

### *5.3.7. Cell culture and transfection studies*

HEK-293T cells were obtained from the American Type Culture Collection and grown at 37°C and 5% $CO_2$ in Dulbecco's Modified Eagle's Medium (DMEM, Gibco). This medium was supplemented with 100 µg/ml penicillin, 100 µg/ml streptomycin and 10% heat-inactivated fetal calf serum (FCS). For dot blot and western blot assays, HEK cells were seeded in a 6-well plate and transfected with the GeneJuice transfection reagent (Merck Millipore). For the coactivation studies, HEK cells were transfected with a mixture of 100 ng luciferase reporter plasmid (preceded by four copies of an ARE), 10 ng AR expression vector, 10 ng TET1 and 10 ng pCMV-β-gal expression plasmid, which served as an internal control for transfection efficiency. On day 1, cells were seeded in a 96-well plate in DMEM containing 5% dextran-coated charcoal-stripped fetal calf serum. The next day, the cells were transfected with the GeneJuice transfection reagent. On day three, the medium was replaced and cells were stimulated with vehicle or with 10 nM of the synthetic ligand methyltrienolone (R1881, Perkin-Elmer, Life Sciences). After 24 hours of stimulation, the cells were harvested in Passive Lysis Buffer (Promega). Luciferase and β-galactosidase activity was measured as described before (Haelens *et al.* 2001). The relative luciferase activities represent the amount of chemiluminescence corrected for the transfection efficiency by normalizing against the β-galactosidase activity. The values shown are the averages of at least three independent experiments performed in triplicate.

### *5.3.8. (h)MEDIP-SEQ*

Three µg of genomic DNA was fragmented to 100-500 bp using the Diagenode bioruptor, end repaired and A-tailed. Following a purification step with dynabeads (Dynal Biotech), sequencing adapters were ligated. For the enrichment step of the methylated DNA immunoprecipitation, 5 µl of an anti-5-methylcytosine antibody or 3 µl of an anti-5-hydroxymethylcytosine antibody coupled to magnetic beads were used. Antibody binding was conducted by overnight incubation in IP buffer (1x PBS, 0.5% BSA, 0.05% Triton X-100). After two hours of incubation with protein A/G magnetic beads (Pierce), the captured samples were washed and bound DNA was eluted by heating the beads for 10 minutes at 99°C. The eluted DNA was amplified by 14 cycles of PCR using Illumina primers, purified and the quality was checked using Picogreen, Nanodrop and BioAnalyzer. Sequencing was performed on the HiSeq2000 (Illumina) as 50 bp, single-end reads. Sequence reads were aligned to the human reference genome (hg19) using Bowtie with –m 1 --strata –best as parameters (this means that only

1 mismatch is allowed, uniquely mapped reads are retained and only the alignment with the best score is reported in case of multiple alignment) (Langmead *et al.* 2009). Model-based Analysis of ChIP-Seq (MACS) was used as the peak-finding algorithm using input DNA as a background control, and after *in silico* extension of the reads by 100 bp (Zhang *et al.* 2008). The entire DIP-Seq protocol and the analyses were performed at the Vesalius Research Center (Leuven, Belgium). GO term and KEGG pathway analyses were performed by the database for annotation, visualization, and integrated discovery (DAVID) programs (Huang da *et al.* 2009a, Huang da *et al.* 2009b).

### *5.3.9. Western blot*

From a cell pellet, whole cell extracts were prepared in HEPES extraction buffer (25 mM HEPES, 300 mM NaCl, 1.5 mM $MgCl_2$, 20 mM β-glycerolphosphate, 2 mM EDTA, 2 mM EGTA, 1 mM DTT, 1% Triton X-100, 10% glycerol, 1 mM $Na_3VO_4$, 50 mM NaF and protease inhibitors) by three freeze-thaw cycli. Fifty µg extract per lane was loaded on NuPage Novex 4-12% Bis-Tris Gels (Life Technologies) and blotted onto PVDF membranes (GE Healthcare). The membranes were probed with the appropriate primary and secondary antibodies (TET1, 1:100; GAPDH, 1:200; both from Santa Cruz Biotechnology), and immunoreactive proteins were visualized by enhanced chemiluminescence (Western Lightning Plus-ECL, Perkin Elmer).

### *5.3.10. Dot blot*

Genomic DNA was isolated using the GenElute Mammalian Genomic DNA Miniprep kit (Sigma-Aldrich). DNA containing 5C, 5mC and 5hmC (Active Motif) were used as controls. Serial dilutions of the DNA were prepared in denaturation buffer (0.4 M NaOH and 10 mM EDTA) and denatured for 10 minutes at 99°. The samples were chilled on ice and neutralized with an equal volume of ice-cold 2 M ammonium acetate. Meanwhile, the membrane was activated in distilled water (Zeta probe blotting membrane, Biorad). After sample loading, the membrane was washed with 0.4 M NaOH and 2x SSC. The DNA was then UV cross-linked to the membrane (UV Light Stratalinker 1800). The membrane was blocked for 1 hour in TBST (Tris-buffered saline containing 0.1% Tween 20) and then probed with the appropriate primary and secondary antibodies (5hmC, 1/1000; 5mC, 1/2000). Modified bases were visualized by enhanced chemiluminescence (Western Lightning Plus-ECL, Perkin Elmer).

### *5.3.11. Immunohistochemistry*

Paraffin-embedded formalin-fixed (FFPE) sections were used for immunohistochemical stainings on 33 PCa samples. Antigen recovery was performed in citrate buffer pH 6 for 20 minutes, followed by incubation with the primary ERG rabbit monoclonal antibody (Abcam), used in a 1:100 dilution. Incubation with the secondary antibody, DAB staining and hematoxylin counterstaining were performed with the BOND polymer refine detection system (Leica). Immunohistochemical stainings were performed on the Bond Max Autostainer (Leica) in collaboration with the Pathology Department of UZ Leuven.

## 5.4. RESULTS

### *5.4.1. Leuven high-risk cohort*

Our cohort consists of 34 patients diagnosed with high-risk PCa and treated with radical prostatectomy. Biological samples were obtained within the frame of the PEARL consortium (ProstatE cAncer Research team Leuven). Tumor samples with at least 50% tumor content were used for copy number profiling, while samples containing at least 75% tumor content were used for both copy number profiling and exome sequencing. Tumor content was estimated by one pathologist (Prof. E. Lerut, Pathology department UZ Leuven). Clinical characteristics of the patients can be found in Supplementary table 5.1, while Supplementary table 5.2. gives an overview of which samples have been used for the different experiments that were performed.

### *5.4.2. Copy number profiling*

#### *5.4.2.1. Detection of amplified and deleted regions*

Analysis of the copy number changes of 32 patients was performed with a combination of the ASCAT and GISTIC software. Four samples failed to give any results, presumably due to low tumor percentage. Another sample could be analyzed, but did not present with any copy number alteration. A representative example of a tumor genome for which copy numbers were determined by ASCAT is shown in Supplementary figure 5.1. We identified regions of recurrent copy number alterations using GISTIC, revealing 12 amplifications and 14 deletions (Figure 5.2). The loss of 8p (containing NKX3.1) and gain of 8q (containing MYC) were recurrent. Other peaks of deletion targeting PTEN (10q23.31) and RB1 (13q14.2) were identified. The deletion of TMPRSS2 (21q22.2) resulting in the fusion of ERG with TMPRSS2 was detected in 11 of the 28 samples. An amplified peak targeting NKX6.2 was also

detected on 10q26.3. All of our recurrently deleted regions have been described before in PCa, while we detected one amplified region that has not been described before in PCa: 7p22.3. It remains to be defined which of the genes contained within this region might be causal oncogenes.

The TMPRSS2-ERG fusion is by far the most common rearrangement in any neoplasm, since it has been detected in almost 50% of all PCa cases examined (Perner *et al.* 2007). Initially, GISTIC analysis was performed on the whole cohort of 28 tumors, but subsequently a GISTIC analysis was also performed on the subgoups. The classification was based on a deletion of the TMPRSS2 gene as determined from the copy number data and ERG overexpression as determined with immunohistochemical staining. We performed a subgroup analysis of 17 patients without TMPRSS2 deletion and 11 patients with deletion of TMPRSS2. We did not find differences in the number of copy number alterations when comparing both groups (data not shown).

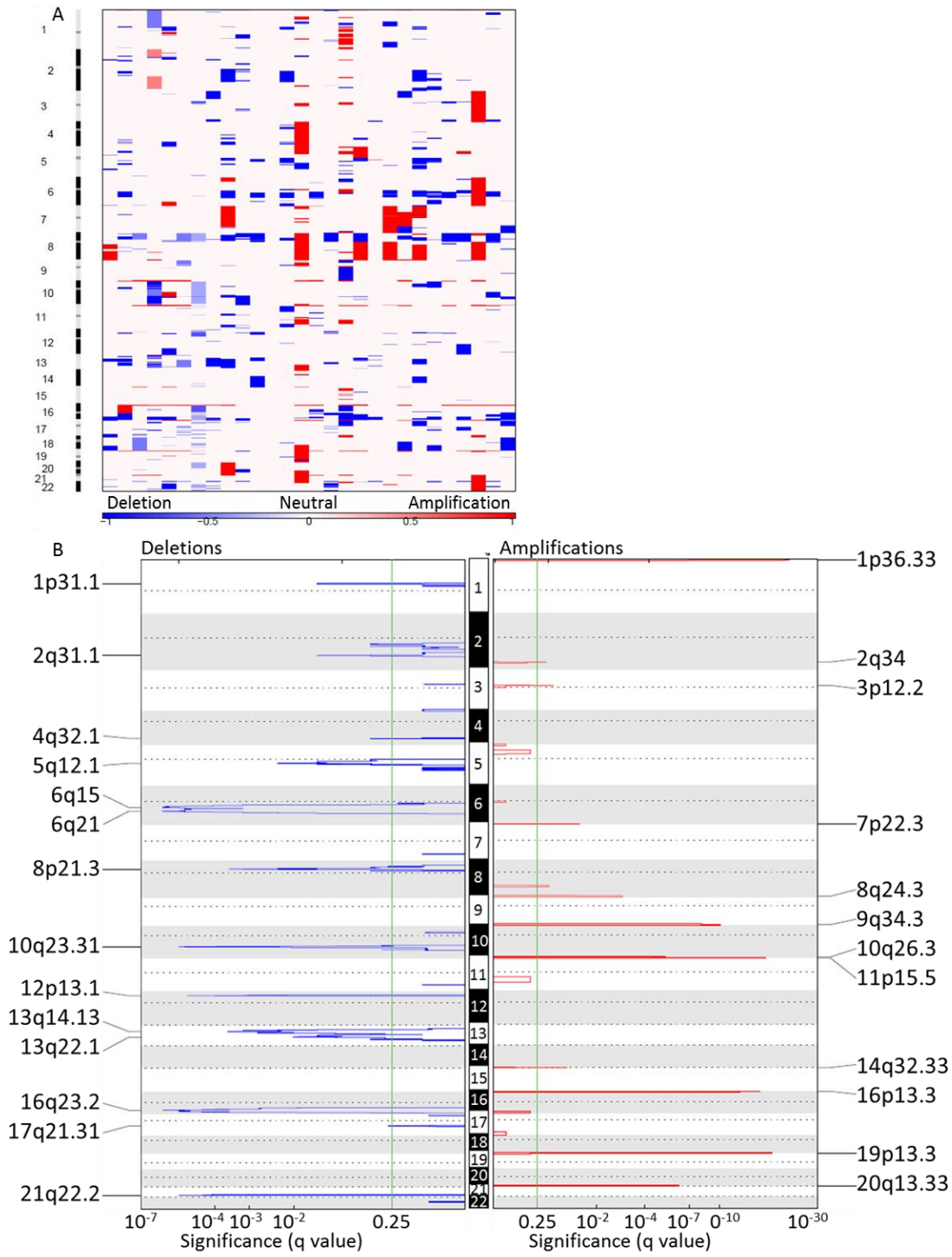



**Figure 5.2.** Significant copy number alterations in the genomes of high-risk PCa. **A.** Amplifications (red) and deletions (blue) are displayed across the genome (chromosome positions, indicated along the y axis, are proportional to marker density) of 28 PCa samples. Each column represents a single tumor sample. **B.** GISTIC analysis of copy number changes in PCa. The statistical significance of the aberrations identified in **A** are displayed as false-discovery rate q values to account for multiple-hypothesis testing. Chromosome positions are indicated along the y axis. Twelve amplifications and 14 deletions surpass the significance threshold (green line). The locations of the peak regions are indicated next to each panel. Processing and primary data extraction were performed at the Vesalius Research Center.

*5.4.2.2. Detection of ERG overexpression using immunohistochemistry*

It has been described before that the TMPRSS2-ERG fusion can be detected with immunohistochemistry instead of with FISH (Park *et al.* 2010). The assay of Park and colleagues had 95.7% sensitivity and 96.5% specificity for detecting the ERG gene fusion in PCa with immunohistochemistry. We wanted to investigate whether this immunoassay also works in our hands, and whether it correlates with the findings of the copy number profiling. Copy number analyses detected a deletion of TMPRSS2 in 11 out of 28 patients. These 11 were confirmed using immunohistochemistry (Figure 5.3). However, in a total of 33 patients, we detected an additional 14 patients of which (part of) the tumor presented with ERG overexpression.

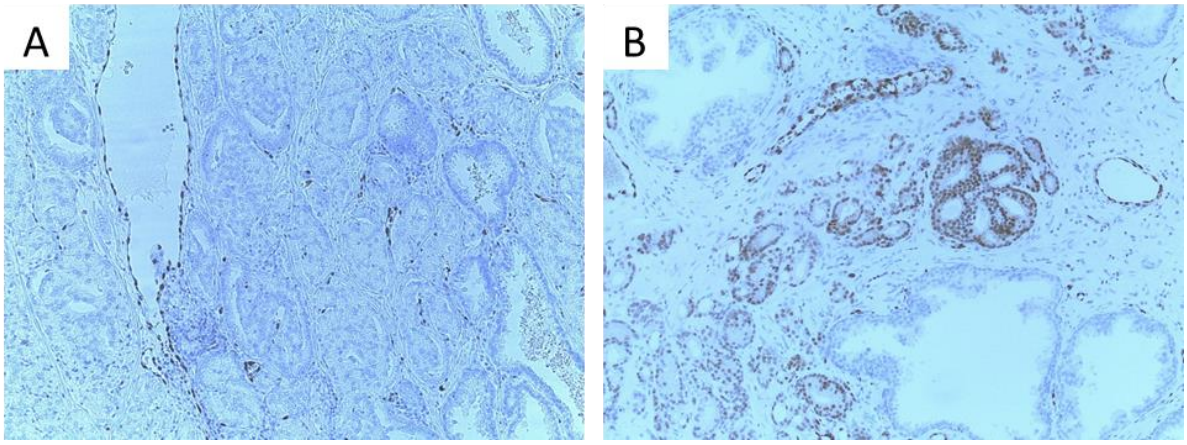



**Figure 5.3.** Representative picture of ERG immunohistochemical staining. The prostate contained tumor foci with and without ERG overexpression. **A.** Tumor focus without ERG overexpression. **B.** Tumor focus with ERG overexpression. Non-tumoral tissue without ERG overexpression can be observed in the upper left and lower right corner. Endothelial cells were used as internal positive control. Images of sample 88824362 were taken at 10x objective magnification.

### *5.4.3. Exome sequencing*

*5.4.3.1. Detection of point mutations using exome sequencing*

Whole exome sequencing using 100 bp paired-end reads was performed on 27 samples of which the tumor content was at least 75% as estimated by the pathologist. On average, 118 million reads were sequenced per sample, with 86% of target bases covered at a depth of ≥20x (Supplementary table 5.3). For variant calling, we used a combination of MuTect and SomaticSniper. We then focused on missense and nonsense mutations that are not present in dbSNP132. Finally, we only retained those mutations that are present in at least 10% of tumor reads, and in less than 2% of non-tumor reads. This led to the identification of 947 somatic mutations that were present in DNA from tumors but absent in DNA from peripheral blood or non-cancerous prostate tissue (Supplementary table 5.4). Of

these, 451 occurred in a single tumor. We hypothesize that a mutation in the Replication Factor C (RFC1) gene could impair DNA repair, resulting in a hypermutated phenotype. Excluding this highly mutated sample, we detected 496 point mutations in 442 different genes. The remaining tumors harbored a median of 19 missense or nonsense mutations (range of 1 to 37 mutations) (Figure 5.4). In one of the 27 samples, we could not detect any point mutation, nor could we detect any change in copy number. We thus propose that this sample either contained a too low amount of tumor cells, or that the tumorigenic process is regulated by epigenetic and translational changes.

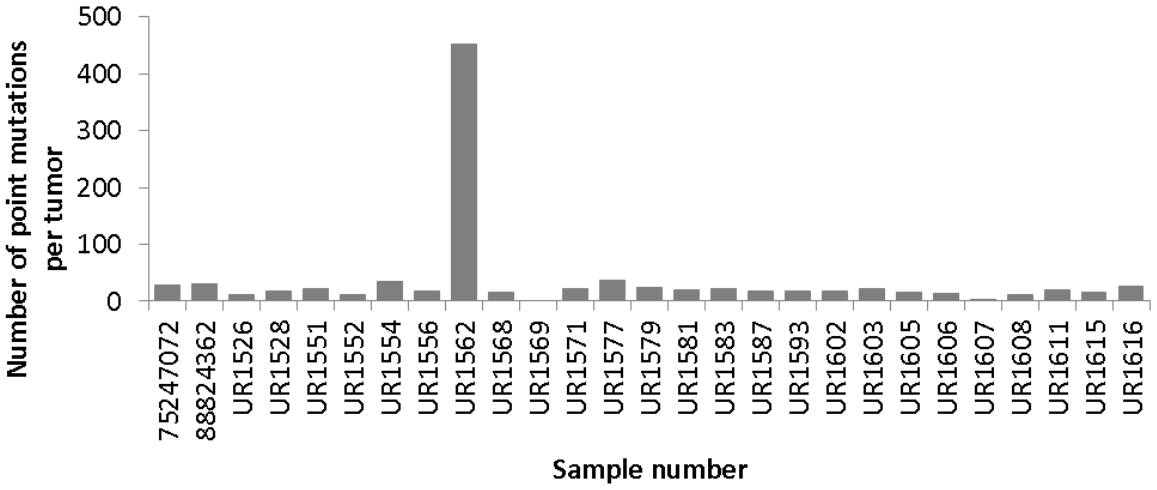



**Figure 5.4.** Histogram for 27 patient samples showing the number of point mutations detected with exome sequencing. One hypermutated samples contains 451 mutations (UR1562), while all the other samples contain 19 mutations on average.

*5.4.3.2. Validation of mutations*

As we used both the MuTect and the SomaticSniper algorithm to call variants, our final list contains mutations that are detected either with both algorithms, or with only one algorithm. A combination of Sanger sequencing and mass spectrometry genotyping was used to validate the detected mutations. To ensure a somatic origin of the variants, both germline and tumor DNA were investigated. A PCR-based amplification of the exonic regions containing the mutations followed by conventional Sanger sequencing was performed for 40 mutations, while mass spectrometry evaluated 314 mutations. In total, 354 mutations were analyzed and 79.4% of variants were validated.

*5.4.3.3. Recurrently mutated genes*

We detected 947 mutations in 831 different genes. Of these genes, seventeen are recurrently mutated in our high-risk PCa cohort (Figure 5.5). It should be noted that the majority of these recurrently mutated genes still needs validation with Sequenom or Sanger sequencing.

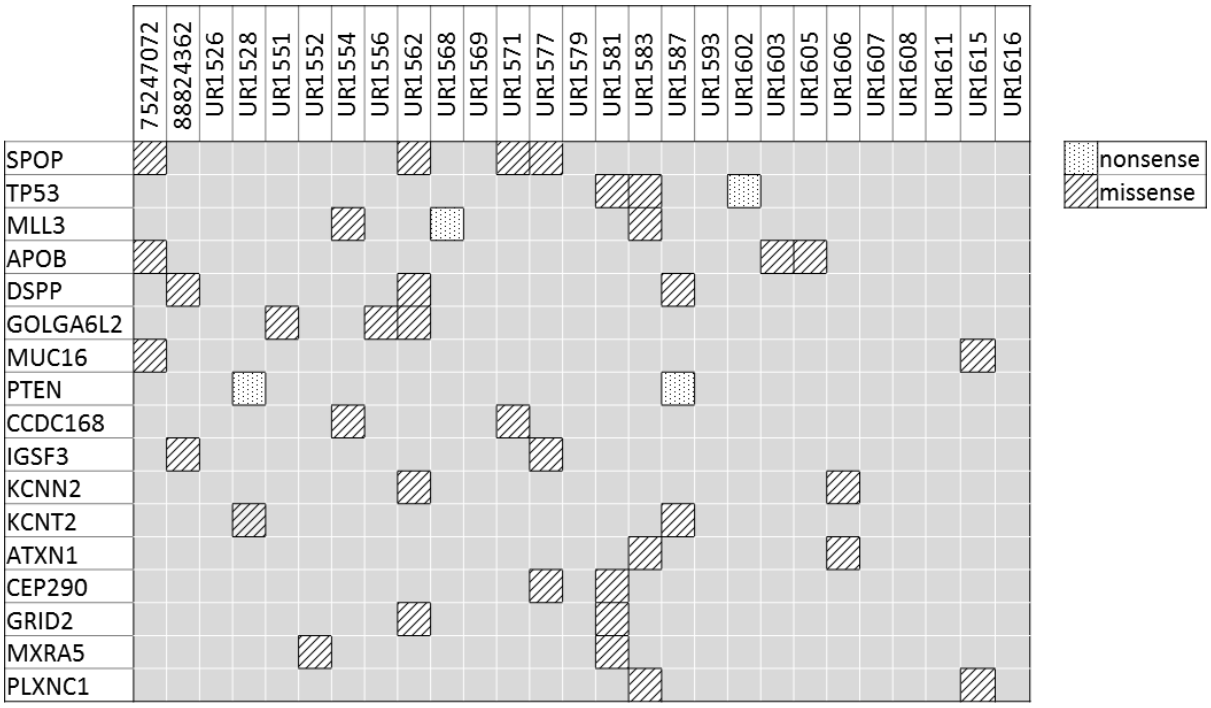



**Figure 5.5.** Representation of recurrently mutated genes, colored by the coding consequence of the mutation. Each column represents a tumor, and each row represents a gene.

### *5.4.4. Mutation in TET1 causes partial loss of function*

In the exome of one of our patients, we detected a A1908S mutation in the TET1 gene. Knowing that DNA methylation is perturbed in almost all cancers and that TET1 is involved in the demethylation process, we wanted to investigate whether this mutation, which is localized in the catalytic domain of TET1, would have a detrimental effect on the enzymatic activity of TET1 (Figure 5.6.A).

*5.4.4.1. Reduced activity of overexpressed mutant TET1*

We first performed a dot blot assay for 5hmC on DNA derived from tumoral and non-tumoral tissue of the patient with the A1908S mutation in the TET1 gene. Indeed, the tumoral tissue presented with a decreased amount of 5hmC compared to non-tumoral tissue (Figure 5.6.B). However, loss of 5hmC is detected in many solid tumors and can be attributed to different causes. We therefore specifically wanted to test the enzymatic activity of A1908S TET1. We overexpressed the catalytic domain of wild

type TET1 in HEK cells, as well as the catalytically inactive mutant (mHxD, H1672Y and D1674A), the A1908S mutant, and the vector only control (Figure 5.6.C). We observed a global increase of 5hmC levels in TET1 overexpressing cells, but not in the mHxD TET1 overexpressing cells. For the A1908S mutant, 5hmC is still formed, but to a lesser extent than in wild type TET1 overexpressing cells. Conversely, 5mC accumulates when mHxD TET1 is overexpressed, while no 5mC remains when wild type TET1 is overexpressed. Again, the level of 5mC present in the A1908S TET1 overexpressing cells, lies somewhere between that of the wild type and that of the mHxD mutant TET1. The overexpression of TET1 was verified by western blot and a similar level of expression for the different TET1 constructs was detected (Figure 5.6.D). Together, these data suggest that the decrease of 5hmC in this patient's tumor is attributable to a decrease in TET1 activity, due to the presence of the A1908S mutation.

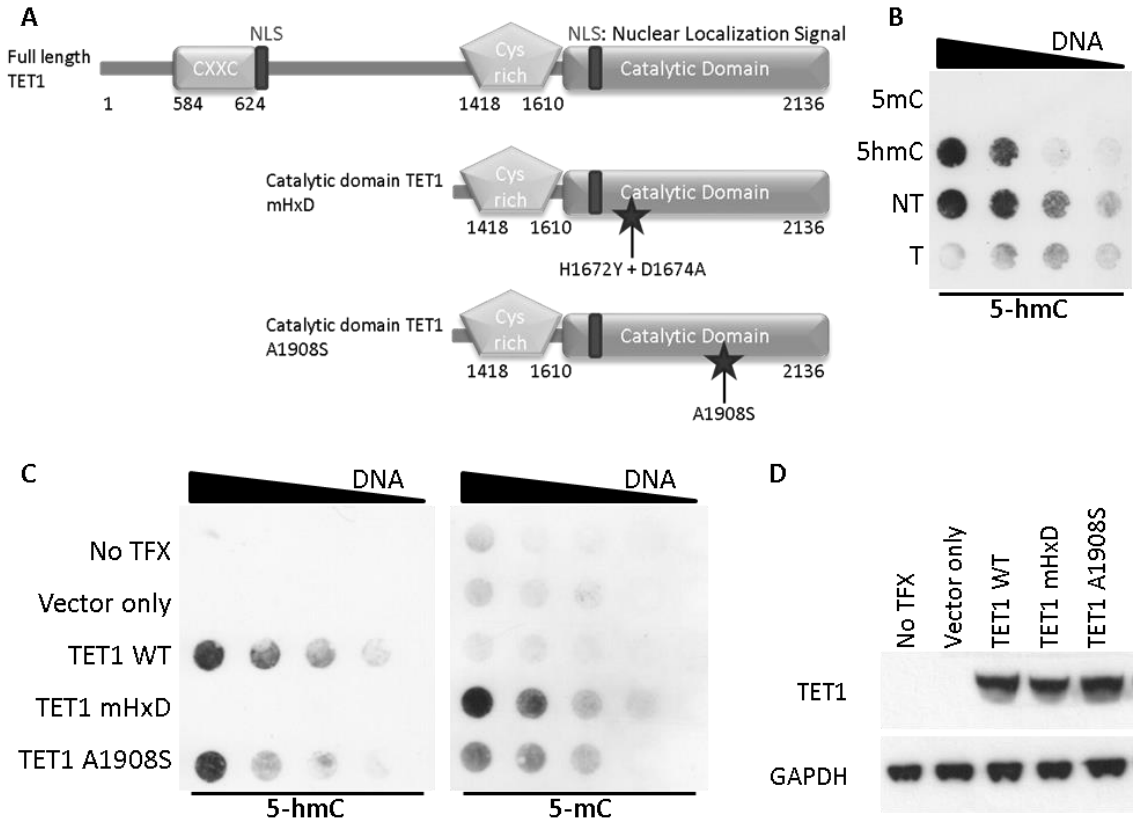



**Figure 5.6.** Overexpression of the A1908S TET1 mutation demonstrates partial loss of the TET1 enzymatic activity. **A.** Schematic diagram of full length TET1 and the catalytic domain of TET1 with the mHxD and A1908S mutations. **B.** Dot blotting shows that the 5hmC level is decreased in tumoral tissue. Genomic DNA isolated from tumoral and non-tumoral tissue of the patient with the TET1 mutation was serially diluted. DNA that contains only 5mC or 5hmC was used as negative and positive control respectively. **C.** Dot blot assay with serial dilutions of DNA, demonstrating global 5mC and 5hmC levels in HEK cells that overexpress the empty control vector, wild type TET1, catalytically dead mHxD mutant or the A1908S mutation. **D.** The expression of TET1 protein by western blot. The same conditions were used as in the dot blot experiment. GAPDH was used as loading control.

*5.4.4.2. Genome-wide mapping of 5mC and 5hmC*

We next asked whether the 5hmC loss in the patient's tumor as determined by dot blot assay was genome-wide or locus-specific. We investigated the 5hmC and 5mC level changes by mapping the genome-wide 5hmC and 5mC distribution using a hydroxymethylated DNA immunoprecipitation (hMeDIP) or a methylated DNA immunoprecipitation (MeDIP) approach coupled with deep sequencing (hMeDIP-Seq and MeDIP-Seq). We compared DNA from the patient's tumoral and non-tumoral tissue with DNA from two other PCa patients without mutations in the TET1 gene. The sample with the TET1 mutation also has a deletion of the PTEN gene as determined by copy number changes and ERG overexpression as determined with immunohistochemistry. The two other samples do not have any of the common features in PCa (such as PTEN deletion, SPOP or TP53 mutation) except for ERG overexpression. Compared to the two tumor samples with wild type TET1, the mutated sample displayed the strongest overall hypo-hydroxymethylation. The present results strongly provide evidence that the 5hmC loss is a genome-wide event and that the TET1 mutation had an effect in the affected cancer tissue (Figure 5.7). Because 5hmC is converted from 5mC by TET enzymes, we reasoned that the decreased 5hmC generation would result in the accumulation of its substrate, 5mC. Indeed, we observed a relative hypermethylation at specific loci. As exemplified in Supplementary figure 5.2, KLK2 and SLC45A3 genes show decreased 5hmC in gene bodies in the tumor sample with mutated TET1 compared with the non-tumoral prostate of the same patient.

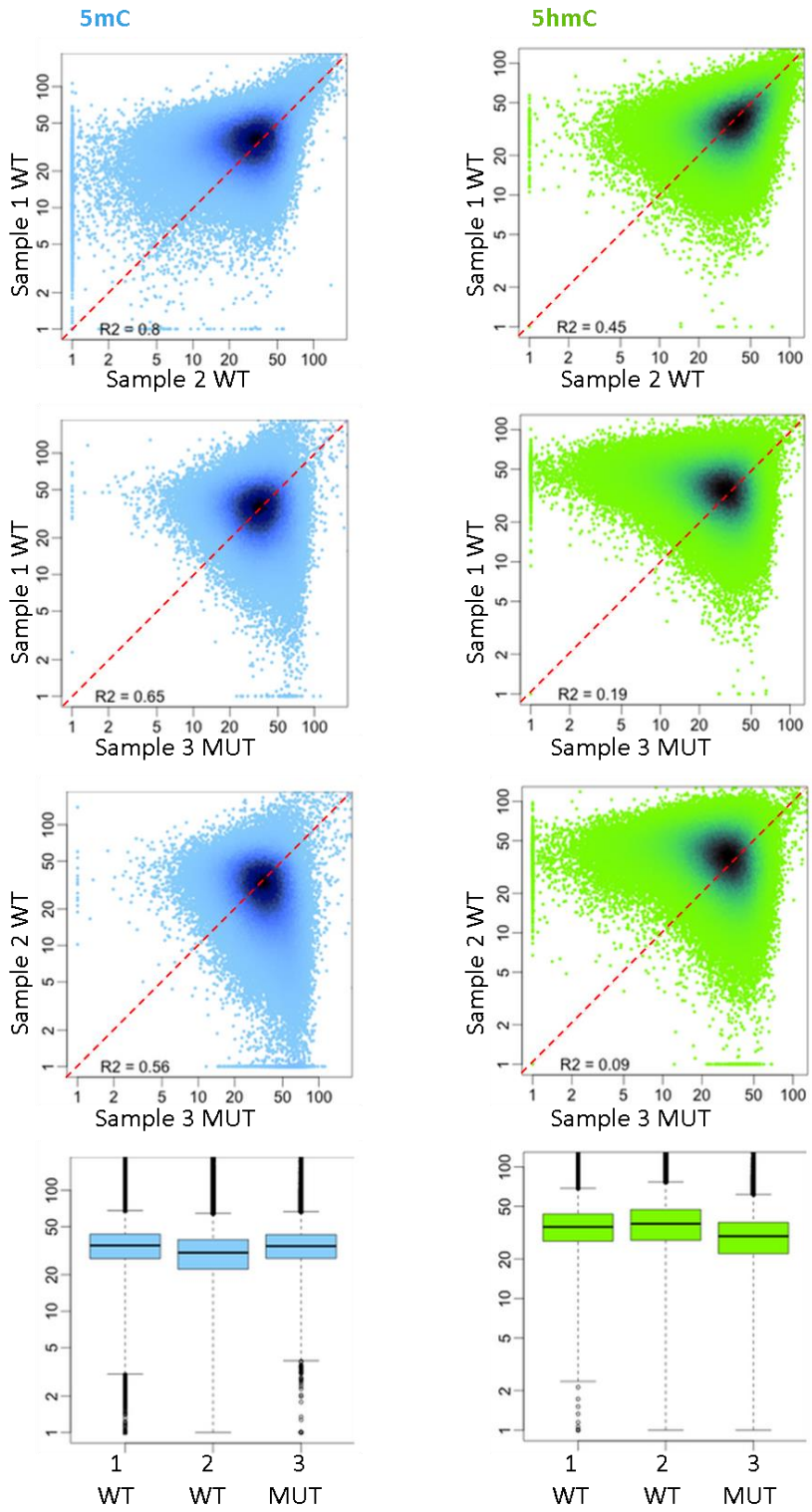



**Figure 5.7.** Genome-wide results for MeDIP-Seq en hMeDIP-Seq. Immunoprecipitation of methylated and hydroxymethylated DNA followed by deep sequencing was performed on DNA isolated from the tumoral and non-tumoral tissues of three patients with PCa. One patient (Sample 3 MUT, 88824362) had the A1908S mutation in TET1, while the other 2 patients (UR1554 and UR1568) did not have a mutation in TET1. An overall decrease in 5hmC and an increase in 5mC at specific loci was observed in the tumor with mutated TET1. The protocol and analyses were performed at the Vesalius Research Center.

*5.4.4.3. Interplay between androgens and 5hmC*

By manually reviewing the hMeDIP-Seq data of the patient with the A1908S TET1 mutation, we noticed that there seemed to be a drastic decrease in 5hmC in some androgen-regulated genes. To explore this more systematically, we gathered a list of 323 androgen-regulated genes. Each of these genes has been detected in at least 2 independent studies as being androgen-regulated (Rajan *et al.* 2011). As can be observed in Figure 5.8, a substantial subset of gene bodies of these genes are indeed hypo-hydroxymethylated in the tumoral tissue of the patient with the TET1 mutation. We generated a list of 55 androgen-regulated genes in which the hydroxymethylation in the tumor is substantially lower than in the non-tumoral tissue of the same patient, or in the tumoral tissues of the two other patients with wild type TET1. KEGG pathway enrichment analysis for the 55 genes revealed that these genes are closely associated with pathways regulating peptidase and protease activity, positive regulation of transcription and gene expression, and positive regulation of apoptosis and cell death.

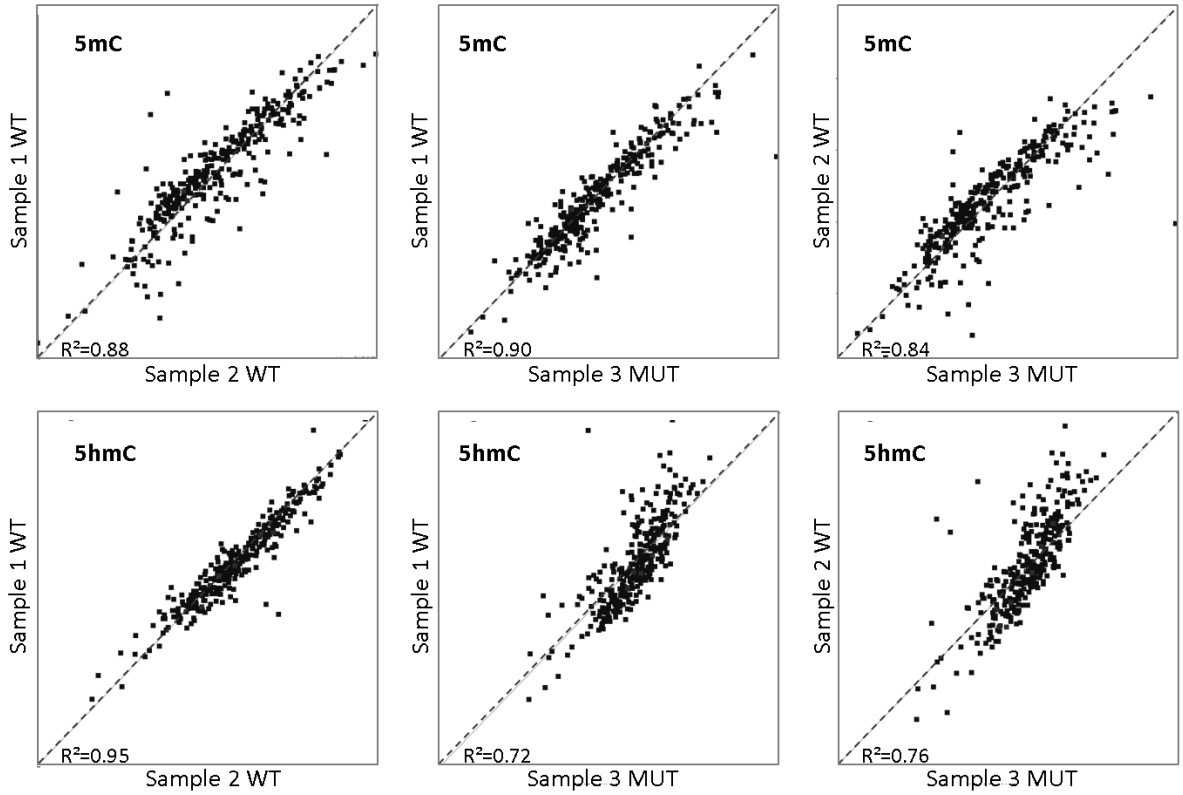



**Figure 5.8.** Genome-wide results for MeDIP-Seq and hMeDIP-Seq of 323 androgen-regulated genes. Immunoprecipitation of methylated and hydroxymethylated DNA followed by deep sequencing was performed on DNA isolated from the tumoral and non-tumoral tissues of three patients with PCa. One patient (Sample 3 MUT, 88824362) had the A1908S mutation in TET1, while the other 2 patients (UR1554 and UR1568) did not have a mutation in TET1.

*5.4.4.4. TET1 is a coactivator of the androgen receptor*

The DIP-Seq experiments showed us that multiple androgen-regulated genes displayed less 5hmC both in the gene body and in the promoter of the sample with the A1908S TET1 mutation. We wanted to investigate whether TET1 can influence the activity of the androgen receptor as a transcription factor. We therefore performed AR coactivation experiments: in the presence of the ligand R1881, an active AR binds to an androgen response element (ARE) which will then result in the activation of the nearby luciferase reporter. In the absence of R1881, luciferase will not be expressed. The addition of increasing amounts of wild type, full length TET1 resulted in an increasing luciferase activity, indicating that TET1 is indeed a coactivator of the AR (Figure 5.9.A).

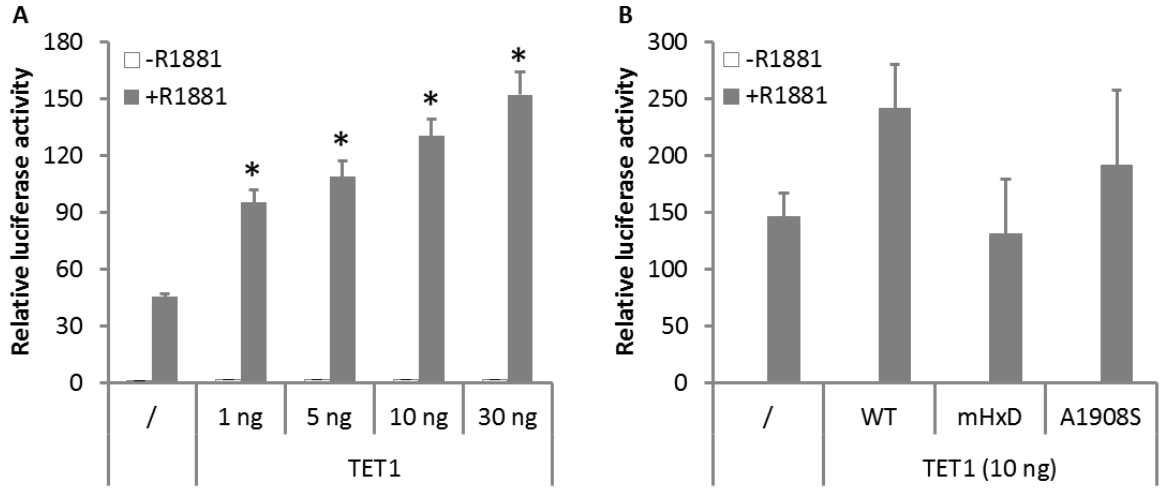



**Figure 5.9.** TET1 is a coactivator of the androgen receptor. **A.** HEK-293T cells were transfected with 100 ng luciferase reporter, 10 ng of AR and pCMV-β-gal each, and increasing amounts of TET1. The next day, cells were incubated in the absence or presence of 10 nM R1881. After 24 hours of stimulation, luciferase and β-galactosidase activities were measured. The relative luciferase activity represents the amount of luminescence corrected for the transfection efficiency by normalizing against β-galactosidase activity. The values shown are the averages of at least 3 independent experiments performed in triplicate. Error bars indicate SEM values. Values that are significantly different compared to the value without TET1 are signalized with a * ($p<0.05$; student's t-test). **B.** Similar experiment as in A, except that 10 ng of different TET1 constructs (wild type, mHxD and A1908S) are transfected in HEK-293T cells.

Knowing that TET1 is a coactivator of the androgen receptor, and that our A1908S mutant showed a decreased enzymatic activity in dot blot assays and hMeDIP-Seq experiments, we wondered whether the mutated TET1 would also result in a decreased coactivation. When the catalytically dead mHxD TET1 is overexpressed, no coactivation is detected (Figure 5.9.B). The overexpression of A1908S TET1 resulted in a tendency to a decreased coactivation compared with the activity of the wild type TET1.

## 5.5. DISCUSSION

Barbieri *et al.* reported the first exome analyses of a cohort of 112 primary PCa samples (Barbieri *et al.* 2012). However, this cohort was very heterogeneous in tumor staging, Gleason grading and predicted risk of recurrence. Here, we sequence a more homogeneous cohort of 27 patients that are predicted to have a high risk of recurrence, as predicted by the d'Amico risk group (D'Amico *et al.* 1998).

### *5.5.1. Copy number profiling*

#### *5.5.1.1. Copy number changes detected in high-risk prostate cancer*

Using SNP arrays, we identified several recurrently amplified and deleted regions in the genomes of primary high-risk prostate cancers. Most of the amplifications and deletions have been described before for the combination of low, intermediate and high-risk PCa (Barbieri *et al.* 2012, Taylor *et al.* 2010). However, we detected a novel amplification of the 7p22.3 region which hasn't been described before. It is possible that one or more oncogenes within this region play an important role specifically in high-risk PCa. To be able to pinpoint the possible oncogenes in these regions, larger cohorts are needed. As the number of analyzed samples increases, the regions will become smaller, and it will become easier to identify the causal oncogenes.

#### *5.5.1.2. TMPRSS2-ERG fusion as detected by copy number and ERG overexpression*

Discordances between the TMPRSS2-ERG fusion status as derived from copy number profiling and ERG overexpression can have different reasons. First of all, in the case of TMPRSS2-ERG, the rearrangement occurs either by a 3 megabase interstitial deletion on a single copy of chromosome 21, or by a chromosomal translocation. Copy number analyses will only identify the fusion when formed by a deletion, while ERG overexpression as detected by immunohistochemistry can be the result of the fusion formed either by deletion or chromosomal translocation. Secondly, ERG overexpression can result not only from a fusion with TMPRSS2, but can also be fused to other genes like SLC45A3 or NDRG1 (Han *et al.* 2008, Pflueger *et al.* 2009). This will result in overexpression of ERG as detected with immunohistochemistry, while this would be undetected by copy number changes. Finally, the biopsy core used for the copy number analyses probably contained only one tumor focus while the ERG overexpression was assessed on sections containing the entire prostate (with multiple foci). Ideally, FISH data and RT-PCR data should be used to unequivocally determine the presence or absence of the TMPRSS2-ERG fusion, but these assays have not been performed in

this study. Nowadays, FISH is still considered to be the gold standard to identify ERG rearrangement, but recent studies provided convincing evidence that the immunohistochemical staining of ERG can be used as a surrogate for the more difficult FISH (Svensson *et al.* 2014).

### *5.5.2. Exome sequencing data*

#### *5.5.2.1. Recurrently mutated genes*

Exome sequencing was performed on 27 cases of primary high-risk PCa. Several of the recurrent mutations in this study have been detected in other cohorts (Barbieri *et al.* 2012, Lindberg *et al.* 2012). Mutations in SPOP, TP53, and PTEN have been reported to occur in 13%, 6% and 4% of PCa samples respectively (Barbieri *et al.* 2012). Recurring mutations in the MLL3 gene were detected in 8% of PCa patients (Lindberg *et al.* 2012). Some of the recurrent mutations in our list, like APOB have not been reported yet, but need confirmation. Indeed, since they occur in genes that are part of large gene families, they might be explained by misalignments of reads rather than genuine mutations. Surprisingly, about half of the mutated genes we detected (200 out of 442 genes) have not been described in high-risk PCa yet. It seems that the more samples will be sequenced, the more likely rare mutations or non-recurrent mutations will be picked up. This is a strong indication of the genetic heterogeneity of the disease. Moreover, some of our novel recurrently mutated genes could well be genes that distinguish between low-risk and high-risk PCa.

#### *5.5.2.2. Mutation in RFC1 could result in a hypermutated tumor*

On average, each sample contains 19 point mutations. One sample however contained 451 mutations. One of these 451 was a mutation in the RFC1 gene to which several functions have been ascribed. The eukaryotic sliding DNA clamp, proliferating cell nuclear antigen (PCNA), is essential for DNA repair synthesis. In order to load PCNA onto the DNA double helix, the ATPase activity of the RFC clamp loader complex is required (Hashiguchi *et al.* 2007). This complex consists of five subunits, of which RFC1 is the large subunit. RFC1 is dispensable for recruitment of PCNA to damaged DNA, yet it is required for the subsequent recruitment of DNA polymerase delta to PCNA. This indicates that RFC1 is essential to stably load the polymerase clamp to enable the start of DNA repair synthesis. The detected mutation in RFC1 could result in an impaired loading of DNA polymerase, leading to impaired DNA repair synthesis and resulting in the almost 500 point mutations. Further studies are required to test this hypothesis.

*5.5.2.3. Molecular classification of prostate cancer*

Large scale integration of copy number aberrations and mutations derived from systematic sequencing studies have led to the definition of molecular subclasses of PCa (Barbieri *et al.* 2013). This classification might signal the start of the transition of PCa from a poorly understood and clinically very heterogeneous disease, to a collection of homogeneous subtypes that are identified by molecular features (Figure 5.10.A). Each genomic subclass might have a distinct prognosis, and hopefully at some point will be treatable with specific targeted therapies. So far, the classification is based on the following features:

- The TMPRSS2-ERG fusion which is present in about half of all PCa patients

- Inactivation of the tumor suppressors PTEN and/or TP53 achieved through point mutations or copy number deletions (in about 30%)

- SPOP mutations, frequently co-occurring with deletion of the CHD1 gene.

A more detailed classification is needed: in our cohort, 9 out of 27 patients cannot be classified into any of these subtypes (Figure 5.10.B). This suggests that the disease is genetically heterogeneous. Maybe these 9 cases should be put in one class of tumors having infrequently mutated genes and the patients might need personalized treatments at later stages of the disease. Indeed, we and others discovered many unique mutations in genes that might be part of signal transduction pathways that are druggable targets.

Moreover, it is also important to study the individual mutations. Indeed, in the case of TET1, a C1746F mutation was reported by Lindberg and colleagues (Lindberg *et al.* 2012). In our dot blot assays and western blots, this mutant had no effect on the enzymatic activity of TET1 (data not shown).

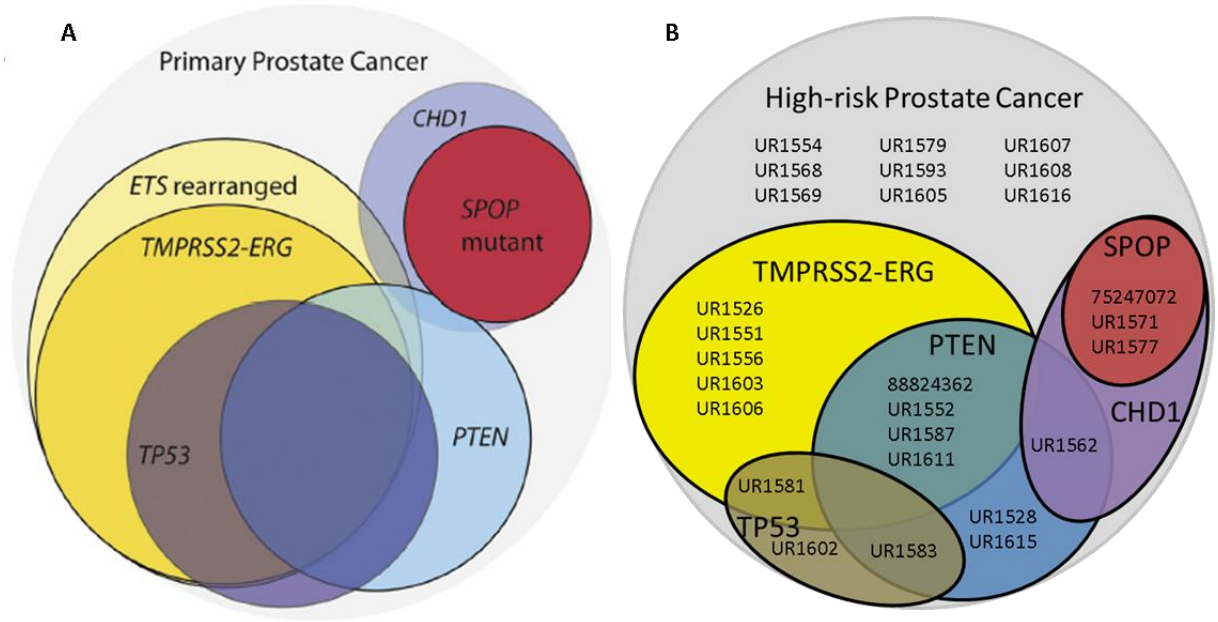



**Figure 5.10.** Molecular classification of high-risk PCa. **A.** About half of all primary PCas harbor TMPRSS2-ERG fusions. The PTEN and TP53 tumor suppressors are deleted or mutated in 20-40% of primary PCa, with significant overlap with each other and with the fusion. SPOP mutations occur in about 10% of all PCas and are mutually exclusive with the fusion, while they are associated with deletions of CHD1. Picture taken from (Barbieri *et al.* 2013). **B.** Same classification as in A, but applied to our 27 high-risk PCa samples. TMPRSS2-ERG means that a deletion of TMPRSS2 was detected using copy number analyses.

### *5.5.2.4. Future perspectives for the high-risk Leuven cohort*

Our current cohort of 34 patients is constantly expanding, with new patients included every month. This however also means that we do not have long-term follow-up or outcome of individual patients yet. Nevertheless, we still feel that our cohort is informative, as it is homogeneous and only contains patients that are predicted to have a high risk of recurrence. Having identified mutated genes in every patient, we know that the number of mutations per patient in our cohort is similar as compared to cohorts consisting of a mix of low-, intermediate- and high-risk patients. Moreover, we were able to identify interesting, novel mutations that have not been implicated in PCa before. Examples are the A1908S mutation in TET1 and the mutation in the RFC1 gene which is involved in the DNA repair processes.

### *5.5.3. Identification of a mutation in TET1 in prostate cancer*

#### *5.5.3.1. Possible mechanisms for the loss of TET1 activity in prostate cancer*

Mammalian cells express three TET genes: TET1, TET2 and TET3. Mutations in TET2 that impair enzymatic activity have been reported frequently in various myeloid leukemias. For example, around 20% of patients with acute myeloid leukemia have a mutation in the TET2 gene (Shih *et al.* 2012). For PCa however, no mutations in TET1 or TET2 have been explicitly described before. A literature study of next-generation sequencing studies on prostate cancer samples revealed two mutations in TET1. A C1746F mutation in the catalytic domain of TET1 as described by Lindberg *et al.* (see also point 5.5.2.3.) and a P204H mutation in a high-risk PCa patient as described by Baca *et al.* (Baca *et al.* 2013, Lindberg *et al.* 2012). We don't think this mutation has a functional effect, as it is located in the N-terminal domain of TET1. Nevertheless, a profound reduction in 5hmC was observed in 30 PCa samples (Haffner *et al.* 2011). Haffner and colleagues did not define their cohort, neither did they report the actual number of samples that contains a reduction in 5hmC. This almost universal reduction of 5hmC levels in PCa can have different causes. First of all, copy number deletion of the region containing a TET gene could result in a downregulation of TET activity. In our cohort, two samples (UR1551 and UR1611) have a deletion in the 10q21.3 region that contains the TET1 gene. Next, TET1 enzymatic activity can also be inhibited by the reduction of alpha-ketoglutarate and accumulation of oncogenic metabolites such as 2-hydroxyglutarate resulting from mutations in IDH1 or IDH2 enzymes (isocitrate dehydrogenase) (Xu *et al.* 2011). However, no such mutations in the IDH1 or IDH2 genes have been described in PCa. Finally, we hypothesize that a missense mutation or transcriptional inactivation of TET1 gene expression could result in a decreased enzymatic activity. Also changes in the activity of enzymes that methylate cytosine (DNA methyl transferases) and enzymes that remove 5hmC intermediates such as TDG and AID can cause the loss of 5hmC in cancer cells (Cadet *et al.* 2013). Thus, cancer-specific metabolic perturbations can influence 5hmC levels and, consequently, alter the epigenetic makeup of the cell (Figueroa *et al.* 2010, Xu *et al.* 2011). In the present study, we identified the first mutation in TET1 in PCa, and in solid tumors in general, that results in a partial loss of enzymatic activity.

#### *5.5.3.2. A1908S mutation in TET1 decreases its enzymatic activity*

Because the TET1 mutation impaired its enzymatic activity *in vitro* (Figure 5.6), we decided to perform a whole genome MeDIP- and hMeDIP-seq experiment on the remaining DNA purified from the tumor sample. Using (h)MeDIP-Seq, we did not find simple correlations between significant loss

of 5hmC, increases of 5mC and expression of associated genes as insufficient material was present to do transcriptome analyses.

The two samples we choose as controls for the (h)MeDIP-Seq experiment seem to have lower 5mC and higher 5hmC compared to the TET1 mutant sample. As TET1 has been described to be less expressed in 33% of primary PCa, it will be crucial to know the TET1 expression status of all our samples (Hsu *et al.* 2012). Therefore, we are currently optimizing immunohistochemical stainings for TET1, 5hmC and 5mC to unequivocally detect the expression levels in both tumoral and non-tumoral tissue.

#### *5.5.3.3. Future challenges for the TET1 mutation*

Understanding the intricate relationship between regulation of 5hmC and associated gene transcription remains a challenge for future basic research. Furthermore, the interplay between androgens and TET1 needs further investigation. Since transfecting full length TET1 is technically challenging, we need to develop alternative approaches to unequivocally prove that TET1 is indeed a coactivator of the androgen receptor. This remains an important issue because of the crucial role of the AR in PCa development and treatments.

Studies have identified TIMP2, TIMP3 and HOXA9 as downstream target genes of TET1, so this could be used to check the functional consequences of the mutated TET1 (Hsu *et al.* 2012, Sun *et al.* 2013). Moreover, a knockdown of TET1 in LNCaP cells results in more aggressive and more invasive tumors, eventually leading to a more metastatic phenotype. Expressing the mutant versus wild type TET1 in these LNCaP cells could also indicate the functional consequences of the mutation.

In conclusion, we detected a novel A1908S mutation in the TET1 gene which causes a partial loss of enzymatic activity. What the consequences are for this individual tumor in terms of aggressiveness, metastatic capacity or biochemical recurrence, remains to be determined.

## 5.6. REFERENCES


Baca, S. C., Prandi, D., Lawrence, M. S*., et al.* (2013). "Punctuated evolution of prostate cancer genomes." Cell **153**(3): 666-677.

Barbieri, C. E., Baca, S. C., Lawrence, M. S*., et al.* (2012). "Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer." Nat Genet **44**(6): 685-689.

Barbieri, C. E., Bangma, C. H., Bjartell, A*., et al.* (2013). "The mutational landscape of prostate cancer." Eur Urol **64**(4): 567-576.

Baylin, S. B. and Jones, P. A. (2011). "A decade of exploring the cancer epigenome - biological and translational implications." Nat Rev Cancer **11**(10): 726-734.

Beroukhim, R., Getz, G., Nghiemphu, L*., et al.* (2007). "Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma." Proc Natl Acad Sci U S A **104**(50): 20007-20012.

Briganti, A., Joniau, S., Gontero, P*., et al.* (2012). "Identifying the best candidate for radical prostatectomy among patients with high-risk prostate cancer." Eur Urol **61**(3): 584-592.

Cadet, J. and Wagner, J. R. (2014). "TET enzymatic oxidation of 5-methylcytosine, 5-hydroxymethylcytosine and 5-formylcytosine." Mutat Res **764-765**: 18-35.

Cibulskis, K., Lawrence, M. S., Carter, S. L*., et al.* (2013). "Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples." Nat Biotechnol **31**(3): 213-219.

D'Amico, A. V., Whittington, R., Malkowicz, S. B*., et al.* (1998). "Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer." JAMA **280**(11): 969-974.

Denayer, S., Helsen, C., Thorrez, L*., et al.* (2010). "The rules of DNA recognition by the androgen receptor." Mol Endocrinol **24**(5): 898-913.

Ferlay, J., Steliarova-Foucher, E., Lortet-Tieulent, J*., et al.* (2013). "Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012." Eur J Cancer **49**(6): 1374-1403.

Figueroa, M. E., Abdel-Wahab, O., Lu, C*., et al.* (2010). "Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation." Cancer Cell **18**(6): 553-567.

Grasso, C. S., Wu, Y. M., Robinson, D. R*., et al.* (2012). "The mutational landscape of lethal castration-resistant prostate cancer." Nature **487**(7406): 239-243.

Guo, J. U., Su, Y., Zhong, C*., et al.* (2011). "Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain." Cell **145**(3): 423-434.

Haelens, A., Verrijdt, G., Callewaert, L*., et al.* (2001). "Androgen-receptor-specific DNA binding to an element in the first exon of the human secretory component gene." Biochem J **353**(Pt 3): 611-620.

Haffner, M. C., Chaux, A., Meeker, A. K*., et al.* (2011). "Global 5-hydroxymethylcytosine content is significantly reduced in tissue stem/progenitor cell compartments and in human cancers." Oncotarget **2**(8): 627-637.

Han, B., Mehra, R., Dhanasekaran, S. M*., et al.* (2008). "A fluorescence in situ hybridization screen for E26 transformation-specific aberrations: identification of DDX5-ETV4 fusion protein in prostate cancer." Cancer Res **68**(18): 7629-7637.

Hashiguchi, K., Matsumoto, Y. and Yasui, A. (2007). "Recruitment of DNA repair synthesis machinery to sites of DNA damage/repair in living human cells." Nucleic Acids Res **35**(9): 2913-2923.

He, Y. F., Li, B. Z., Li, Z*., et al.* (2011). "Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA." Science **333**(6047): 1303-1307.

Hotchkiss, R. D. (1948). "The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography." J Biol Chem **175**(1): 315-332.

Hsu, C. H., Peng, K. L., Kang, M. L*., et al.* (2012). "TET1 suppresses cancer invasion by activating the tissue inhibitors of metalloproteinases." Cell Rep **2**(3): 568-579.

Huang da, W., Sherman, B. T. and Lempicki, R. A. (2009a). "Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists." Nucleic Acids Res **37**(1): 1-13.

Huang da, W., Sherman, B. T. and Lempicki, R. A. (2009b). "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources." Nat Protoc **4**(1): 44-57.

Ito, S., Shen, L., Dai, Q.*, et al.* (2011). "Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine." Science **333**(6047): 1300-1303.

Jin, S. G., Jiang, Y., Qiu, R.*, et al.* (2011). "5-Hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations." Cancer Res **71**(24): 7360-7365.

Kudo, Y., Tateishi, K., Yamamoto, K.*, et al.* (2012). "Loss of 5-hydroxymethylcytosine is accompanied with malignant cellular transformation." Cancer Sci **103**(4): 670-676.

Langmead, B., Trapnell, C., Pop, M.*, et al.* (2009). "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome." Genome Biol **10**(3): R25.

Larson, D. E., Harris, C. C., Chen, K.*, et al.* (2012). "SomaticSniper: identification of somatic point mutations in whole genome sequencing data." Bioinformatics **28**(3): 311-317.

Li, H. and Durbin, R. (2009). "Fast and accurate short read alignment with Burrows-Wheeler transform." Bioinformatics **25**(14): 1754-1760.

Lindberg, J., Mills, I. G., Klevebring, D.*, et al.* (2013). "The mitochondrial and autosomal mutation landscapes of prostate cancer." Eur Urol **63**(4): 702-708.

McKenna, A., Hanna, M., Banks, E.*, et al.* (2010). "The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data." Genome Res **20**(9): 1297-1303.

Meng, M. V., Elkin, E. P., Latini, D. M.*, et al.* (2005). "Treatment of patients with high risk localized prostate cancer: results from cancer of the prostate strategic urological research endeavor (CaPSURE)." J Urol **173**(5): 1557-1561.

Park, K., Tomlins, S. A., Mudaliar, K. M.*, et al.* (2010). "Antibody-based detection of ERG rearrangement-positive prostate cancer." Neoplasia **12**(7): 590-598.

Perner, S., Mosquera, J. M., Demichelis, F.*, et al.* (2007). "TMPRSS2-ERG fusion prostate cancer: an early molecular event associated with invasion." Am J Surg Pathol **31**(6): 882-888.

Pflueger, D., Rickman, D. S., Sboner, A.*, et al.* (2009). "N-myc downstream regulated gene 1 (NDRG1) is fused to ERG in prostate cancer." Neoplasia **11**(8): 804-811.

Rajan, P., Dalgliesh, C., Carling, P. J.*, et al.* (2011). "Identification of novel androgen-regulated pathways and mRNA isoforms through genome-wide exon-specific profiling of the LNCaP transcriptome." PLoS One **6**(12): e29088.

Shih, A. H., Abdel-Wahab, O., Patel, J. P.*, et al.* (2012). "The role of mutations in epigenetic regulators in myeloid malignancies." Nat Rev Cancer **12**(9): 599-612.

Sun, M., Song, C. X., Huang, H.*, et al.* (2013). "HMGA2/TET1/HOXA9 signaling pathway regulates breast cancer growth and metastasis." Proc Natl Acad Sci U S A **110**(24): 9920-9925.

Svensson, M. A., Perner, S., Ohlson, A. L.*, et al.* (2014). "A Comparative Study of ERG Status Assessment on DNA, mRNA, and Protein Levels Using Unique Samples from a Swedish Biopsy Cohort." Appl Immunohistochem Mol Morphol **22**(2): 136-141.

Tahiliani, M., Koh, K. P., Shen, Y.*, et al.* (2009). "Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1." Science **324**(5929): 930-935.

Taylor, B. S., Schultz, N., Hieronymus, H.*, et al.* (2010). "Integrative genomic profiling of human prostate cancer." Cancer Cell **18**(1): 11-22.

Van Loo, P., Nordgard, S. H., Lingjaerde, O. C.*, et al.* (2010). "Allele-specific copy number analysis of tumors." Proc Natl Acad Sci U S A **107**(39): 16910-16915.

Xu, W., Yang, H., Liu, Y.*, et al.* (2011). "Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases." Cancer Cell **19**(1): 17-30.

Yang, H., Liu, Y., Bai, F.*, et al.* (2013). "Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation." Oncogene **32**(5): 663-669.

Zhang, Y., Liu, T., Meyer, C. A.*, et al.* (2008). "Model-based analysis of ChIP-Seq (MACS)." Genome Biol **9**(9): R137.

## 5.7. SUPPLEMENTARY INFORMATION

**Supplementary table 5.1.** Clinical characteristics of prostate cancer genomes.

**Supplementary table 5.2.** Overview of which samples have been used for certain experiments.

**Supplementary table 5.3.** Summary of alignment results over 27 samples**.**

**Supplementary table 5.4.** List of all 519 detected somatic point mutations. The table is not included in this thesis due to space constraints.

**Supplementary figure 5.1.** A representative example of a tumor genome for which copy numbers were determined by ASCAT.

**Supplementary figure 5.2.** MeDIP-Seq and hMeDIP-Seq results of KLK2 and SLC45A3 genes.

**Supplementary table 5.1.** Clinical characteristics of prostate cancer genomes.

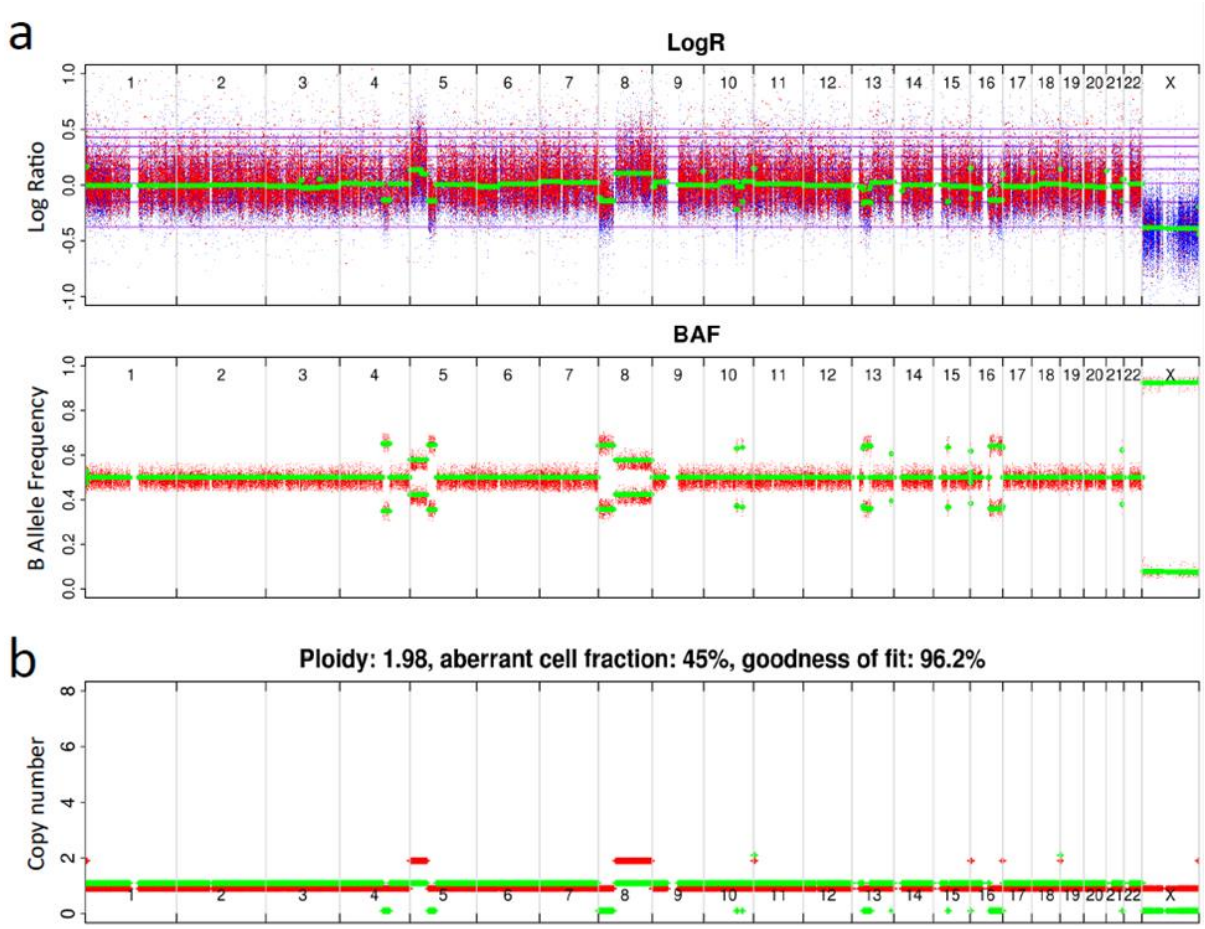
| Sample | Date of RP | Age | Gleason score | Pathological stage | PSA (ng/ml) | Positive surgical margins | Last follow up Date | PSA |
|---|---|---|---|---|---|---|---|---|
| V2490 | 8/11/2011 | 51 | 4+3 | T3aN1 | 24,75 | 0 | 15/01/2014 | 0 |
| 88824362 | 17/11/2011 | 63 | 4+3 | T3aN0 | 10,30 | 0 | 5/12/2013 | 0 |
| 75247072 | 5/12/2011 | 65 | 4+4 | T3aN0 | 7,90 | 0 | 11/10/2013 | 0 |
| UR1526 | 5/06/2012 | 51 | 4+3 | T3bN1 | 27,00 | 0 | 6/01/2014 | 0 |
| UR1528 | 18/06/2012 | 63 | 4+3 | T3aN1 | 10,00 | 0 | 12/12/2013 | 0,03 |
| UR1551 | 17/10/2012 | 62 | 3+4 | T3aN0 | 6,10 | 0 | 7/10/2013 | 0 |
| UR1552 | 29/10/2012 | 56 | 3+4 | T3aN0 | 30,22 | 0 | 12/09/2013 | 0 |
| UR1554 | 5/11/2012 | 65 | 3+4 | T3aN0 | 8,95 | 0 | 18/10/2013 | 0 |
| UR1556 | 12/11/2012 | 64 | 3+4 | T2cN0 | 4,58 | 0 | 12/12/2013 | 0 |
| UR1560 | 3/12/2012 | 71 | 4+4 | T3aN0 | 15,64 | 0 | 29/07/2013 | 0 |
| UR1562 | 13/12/2012 | 74 | 4+5 | T3bN0 | 5,99 | 0 | 18/09/2013 | 0 |
| UR1566 | 17/01/2013 | 69 | 4+3 | T3aN0 | 5,20 | 1 | 6/11/2013 | 0 |
| UR1568 | 21/01/2013 | 62 | 3+4 | T2cN0 | 5,10 | 1 | 2/01/2014 | 0,05 |
| UR1569 | 24/01/2013 | 63 | 4+3 | T2cN0 | 4,66 | 0 | 20/01/2014 | 0 |
| UR1570 | 25/01/2013 | 74 | 4+5 | T3aN1 | 19,00 | 1 | 28/05/2013 | 0,08 |
| UR1571 | 31/01/2013 | 62 | 4+4 | T3bN0 | 28,00 | 1 | 6/01/2014 | 0 |
| UR1577 | 21/03/2013 | 75 | 4+5 | T3bN0 | 10,90 | 0 | 30/01/2014 | 0 |
| UR1579 | 25/03/2013 | 67 | 3+4 | T3aN0 | 6,70 | 1 | 6/01/2014 | 0 |
| UR1581 | 4/04/2013 | 69 | 4+3 | T3aN0 | 18,00 | 1 | 30/01/2014 | 0 |
| UR1583 | 18/04/2013 | 65 | 4+5 | T3bN1 | 15,20 | 1 | 6/01/2014 | 0,72 |
| UR1586 | 30/05/2013 | 67 | 4+3 | T3aN0 | 4,16 | 0 | 10/01/2014 | 0,04 |
| UR1587 | 31/05/2013 | 56 | 4+3 | T2cN0 | 10,60 | 0 | 30/01/2014 | 0 |
| UR1591 | 21/06/2013 | 62 | 3+4 | T2cN0 | 15,00 | 0 | 12/12/2013 | 0,003 |
| UR1593 | 27/06/2013 | 73 | 3+4 | T2cN0 | 9,50 | 0 | 4/12/2013 | 0 |
| UR1602 | 1/08/2013 | 63 | 5+4 | T4N1 | 9,26 | 1 | 8/01/2014 | 0 |
| UR1603 | 12/08/2013 | 67 | 4+3 | T3aN0 | 10,40 | 0 | 3/01/2014 | 0 |
| UR1605 | 29/08/2013 | 73 | 3+4 | T3aN0 | 7,00 | 0 | 23/01/2014 | 0 |
| UR1606 | 30/08/2013 | 73 | 4+3 | T2cN0 | 4,80 | 1 | | |
| UR1607 | 9/09/2013 | 63 | 3+4 | T2cN0 | 2,49 | 0 | 23/12/2013 | 0 |
| UR1608 | 16/09/2013 | 67 | 5+4 | T3bN1 | 8,00 | 0 | 31/10/2013 | 0 |
| UR1611 | 3/10/2013 | 61 | 4+3 | T3aN0 | 1,36 | 0 | 7/11/2013 | 0,07 |
| UR1615 | 31/10/2013 | 68 | 3+4 | T3bN0 | 5,38 | 0 | 19/12/2013 | 0 |
| UR1616 | 4/11/2013 | 55 | 4+4 | T3bN1 | 21,68 | 0 | 19/12/2013 | 0,14 |

**Supplementary table 5.2.** Overview of which samples have been used for certain experiments. Gray and light gray samples have been used for the corresponding experiments, but results could only be obtained for the green samples.

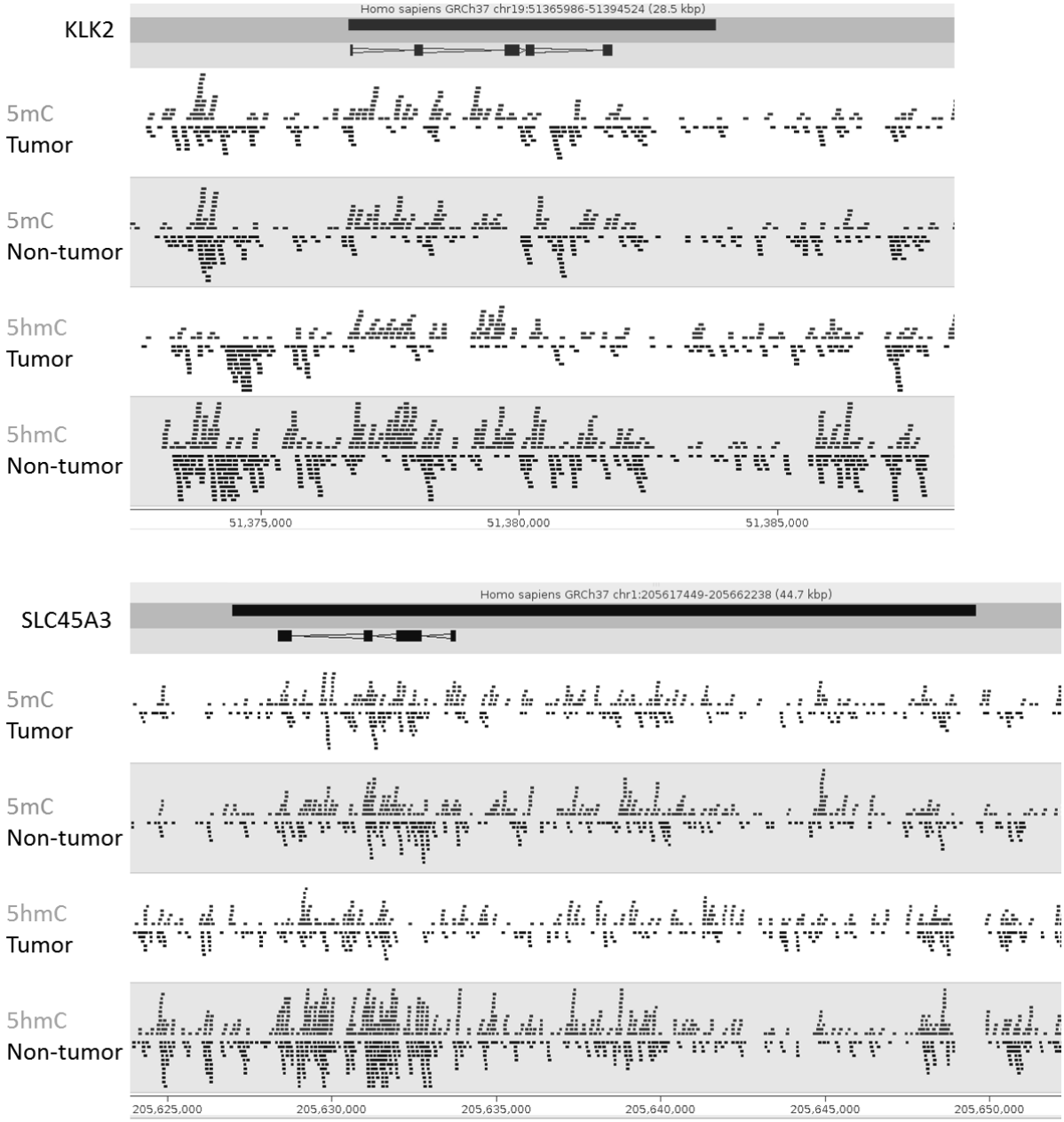
| Copy number alterations | Whole exome sequencing | Sequenom validation | DIP-SEQ | IHC ERG | Sample | Tumor percentage as estimated by pathologist |
|---|---|---|---|---|---|---|
| x | | | | x | V2490 | 100% tumor |
| x | x | x | | x | 75247072 | 100% tumor |
| x | x | x | | x | UR1526 | 90% tumor |
| x | x | x | | x | UR1528 | 65% tumor |
| x | x | | x | x | 88824362 | 100% tumor |
| x | x | x | | x | UR1551 | 75% tumor |
| x | x | x | | x | UR1552 | 100% tumor |
| x | x | x | x | x | UR1554 | 100% tumor |
| x | x | x | | x | UR1556 | 100% tumor |
| x | | | | x | UR1560 | 50% tumor |
| x | x | x | | x | UR1562 | 100% tumor |
| x | | | | x | UR1566 | 50% tumor |
| x | x | x | x | x | UR1568 | 90% tumor |
| x | x | | | x | UR1569 | 95% tumor |
| x | | | | x | UR1570 | 50% tumor |
| x | x | x | | | UR1571 | 75% tumor |
| x | x | x | | x | UR1577 | 75% tumor |
| x | x | x | | x | UR1579 | 95% tumor |
| x | x | x | | x | UR1581 | 100% tumor |
| x | x | x | | x | UR1583 | 95% tumor |
| | | | | x | UR1586 | 50% tumor |
| x | x | x | | x | UR1587 | 90% tumor |
| x | | | | x | UR1591 | 50% tumor |
| x | x | x | | x | UR1593 | 100% tumor |
| x | x | x | | x | UR1602 | 90% tumor |
| x | x | x | | x | UR1603 | 90% tumor |
| x | x | x | | x | UR1605 | 95% tumor |
| x | x | x | | x | UR1606 | 95% tumor |
| x | x | x | | x | UR1607 | 100% tumor |
| x | x | x | | x | UR1608 | 100% tumor |
| x | x | x | | x | UR1611 | 100% tumor |
| x | x | x | | x | UR1615 | 100% tumor |
| x | x | x | | x | UR1616 | 75% tumor |

**Supplementary table 5.3.** Summary of alignment results over 54 samples

| | Average | St. Dev. | Minimum | Maximum |
|---|---|---|---|---|
| Total number of reads (Mbp) | 118,10 | 41,79 | 49,25 | 216,36 |
| Number of high quality aligned bases (Gbp) | 10,38 | 3,79 | 4,30 | 18,95 |
| Mean read length | 101 | 0 | 101 | 101 |
| Percentage targets with zero coverage | 2,55 | 1,47 | 1,81 | 10,00 |
| Percentage of target bases covered at least 2x | 95,84 | 1,95 | 86,15 | 97,07 |
| Percentage of target bases covered at least 10x | 92,43 | 2,64 | 80,23 | 95,08 |
| Percentage of target bases covered at least 20x | 86,10 | 5,09 | 73,34 | 93,37 |
| Percentage of target bases covered at least 30x | 76,10 | 9,99 | 49,18 | 91,31 |




**Supplementary figure 5.1.** A representative example of a tumor genome for which copy numbers were determined by ASCAT. **A.** The LogR track quantifies the copy number of each genomic locus and can therefore be used to distinguish amplifications from deletions. The B-allele frequency or BAF shows the relative presence of each of the two alternative nucleotides at each SNP locus profiled and is used to distinguish which allele is amplified or deleted. **B.** The ASCAT profile contains the allele-specific copy number of all assayed loci: the copy number is shown on the y axis, the genomic location on the x axis. The allele with the lowest copy number is shown in green, the allele with the highest copy number in red. Both lines are slightly shifted such that they do not overlap. Representative example of sample 88824362.




**Supplementary figure 5.2.** MeDIP-Seq and hMeDIP-Seq results of KLK2 and SLC45A3 genes. Figures are generated using the SeqMonk software, where the upper tags represent reads on the forward strand and lower tags reads on the reverse strand. Example for DNA isolated from the prostate of patient 88824362 with the A1908S mutation.

# CHAPTER 6:

# GENERAL DISCUSSION AND PERSPECTIVES

This study exemplifies how next-generation sequencing can contribute to a deeper understanding of the genetic landscape of PCa. We first analyzed two commonly used PCa cell lines (LNCaP and C4-2B) and consequently adopted the acquired knowledge and technological expertise to the analysis of high-risk PCa patient samples. This provided new insights into the genetics of PCa, revealed novel mutations and generated new hypotheses that can be tested in future research.

## 6.1. MODEL SYSTEMS FOR PROSTATE CANCER

### *6.1.1. Cell lines as models for prostate cancer*

In general, cell lines can be beautiful tools to study many aspects of tissue and cancer biology, in as far as they haven't differentiated too much from the tissue they were derived from. This also holds true for PCa. In most cases, PCa is inherently slow growing, and morphologically and molecularly heterogeneous. Therefore, cells from prostate carcinomas are one of the most difficult cell types from which to establish continuous cell lines (van Bokhoven *et al.* 2003). Despite numerous attempts to derive cells from primary tumors, the frequently used model systems for PCa (LNCaP, DU145 and PC-3) were derived from metastatic lesions and either have lost the AR, or have a mutated AR. These immortalized cell lines are therefore not representative of primary disease. Be that as it may, LNCaP cells still express many prostate (cancer) markers, are androgen-dependent in their growth and hence can be used to study androgen transcription and cell cycle control.

Unfortunately, each cell line represents only one tumor type and one cancer patient. Combining this with the heterogeneity of PCa, the reproduction of studies in several cell lines can lead to inconsistent or contradicting results. Nevertheless, cell lines are easily cultured, readily available and simple to use and therefore still used frequently in preclinical research. Novel models are being developed that should close the (preclinical) gap between studies on cell lines and clinical studies.

### *6.1.2. Recent developments in prostate cancer model systems*

More and more evidence is emerging about the crucial interplay between tumor cells and their microenvironment during the development and progression of PCa. The tumor microenvironment denotes the intimate relationship between malignant cells and their surrounding stroma, which

includes fibroblasts, lymphocytes, blood vessels and extracellular matrix components (Centenera *et al.* 2013). The signaling between stromal and epithelial compartments is crucial for PCa development. Cell lines however do not accurately recapitulate the heterogeneity or complexity of this tissue microenvironment.

Recent developments of novel model systems for PCa are the *ex vivo* cultures of biopsies (explants) and the patient-derived xenografts. As both models are derived from primary human tissues and still contain a stromal department, they represent more relevant systems to study PCa. The explant cultures allow taking into account the intrapatient heterogeneity as well as the extent of proliferative responses to steroid hormones (Centenera *et al.* 2012). Furthermore, this model allows the relatively rapid assessment of the antiproliferative activity of novel therapeutic agents in the tissue, without having to treat the individual patients (Centenera *et al.* 2013). Alternatively, xenografts may be derived directly from patient tissue without prior *in vitro* culture. This means that tumor tissue that is excised at the time of surgery is immediately transplanted into immune-deficient mice. As for the *ex vivo* culturing, this model allows testing different treatment options, together with the induction of castration-resistance (Lin *et al.* 2014).

### *6.1.3. LNCaP and C4-2B cells as model system*

Despite recent advances in the development of novel model systems that more closely resemble the situation as it is in the patient, the LNCaP and C4-2B cells are still frequently used as a model for the progression of androgen-dependent to androgen-independent PCa. We therefore set out to catalogue all point mutations in the exomes of both cell lines. One should note that the field of massively parallel sequencing is rapidly evolving. Hence, the bioinformatics analyses and algorithms used in Chapter 3 (Variations in the exome of the LNCaP PCa cell line) and Chapter 4 (Comparative genomic and transcriptomic analyses of LNCaP and C4-2B PCa cell lines) were different. In the LNCaP exome in Chapter 3, the BWA, Bowtie, Samtools and Varscan algorithms, which were available at that time, were used. We reanalyzed this exome using BWA and GATK (Genome Analysis Toolkit) UnifiedGenotyper to be able to compare it with the exome of C4-2B cells (Chapter 4). For LNCaP cells, the differences between both analyses are relatively small, although each method identifies some mutations that are not identified with the other method. Currently, the field has evolved again and the GATK UnifiedGenotyper is now substituted by the HaplotypeCaller (GATK Best Practices).

In both cell lines, we identified a surprisingly high number of mutations (between 1800 and 4000). The filtering used to obtain this number was very stringent in comparison to the filtering applied to primary PCa tissues. Nevertheless, we identified a 100 times more mutations in the cell lines than in

primary PCa. This could be attributed to the DNA mismatch repair deficiency, the metastatic origin of both cell lines as compared to primary tumor samples and the adaptation to the prolonged culturing (Chen *et al.* 2001, Leach *et al.* 2000). Although we have no knowledge as to the effect of each of these cell line mutations, one should be very careful when studying certain genes, as about 8% of all genes are mutated in these cells.

The comparison of exome sequencing and transcriptome analyses of the LNCaP and C4-2B cells can be used to generate or test hypotheses for further research. We looked for an affected pathway that could explain the high metastatic potential of C4-2B and its preference to metastasize to bone, characteristics that are absent in the LNCaP cells. We found that the pathway converging on MLCK (myosin light chain kinase) might be a start for novel functional tests. However, since our primary goal was the sequencing of high-risk PCa and samples became available, we did not investigate this further.

## 6.2. HIGH-RISK PROSTATE CANCER

### *6.2.1. The high-risk Leuven cohort*

A Pan-cancer analysis revealed that genes mutated in over 20% of cancers have largely been discovered. However, only a handful of cancer genes are mutated at that high frequency, and most cancer genes that are mutated in patients occur at intermediate frequencies between 2 and 20% (Lawrence *et al.* 2014). To provide therapeutic options for most patients, it will therefore be critical to identify and understand the pathway-level implications of all genes mutated at intermediate frequencies. This means that, despite the fact that a relatively large number of PCa exomes have already been sequenced, it is still valuable to sequence novel exomes from new patient cohorts as we did in the present study. Specifically, the present study focused on high-risk PCa, resulting in a relatively homogeneous group. Our goal was not only to list novel mutated genes, but also to try to identify new cancer driver mutations by testing the functional consequences of specific mutations. Because the size of our cohort is relatively small (it is still growing), and because it is a prospective study, the generated data are currently mainly hypothesis-generating. It is crucial to first test these hypotheses *in vitro* (in cell lines) and in a retrospective study of the high-risk and lethal PCa samples collected at UZ Leuven, and then return to the clinic for further validation.

### *6.2.2. Planned experimental studies*

We identified a novel point mutation in the TET1 gene in one of the high-risk PCa samples. This mutation results in a partial loss of the DNA hydroxymethylating activity. However, we still need to investigate what the consequences of this mutation could be for tumor growth, tumor aggressiveness, tumor invasion and possible tumor lethality. Early experiments could involve using a PCa cell line that has a low endogenous expression of TET1 and then overexpress either wild type or mutant TET1. Growth and metastatic properties could then be analyzed *in vitro* or in xenotransplant studies. Alternatively, we could knockout the TET1 gene or introduce the mutation in a mouse model and subsequently check for PCa development (after crossing in other PCa mutations like PTEN deletion, the TMPRSS2-ERG fusion etcetera). This could be further correlated with changed TET1 activity in primary and metastatic PCa.

Similar to the TET1 mutation, we are currently investigating a novel point mutation in the eEF2K gene or the eukaryotic elongation factor-2 kinase. This kinase phosphorylates and inactivates eEF2, thus preventing eEF2 from binding to the ribosome and hence blocking protein synthesis (Figure 6.1.A). Many signaling pathways (such as MAPK and mTOR) control eEF2 activity, but this regulation is exerted exclusively via modification of eEF2K activity rather than eEF2 activity (Hizli *et al.* 2013). Increased eEF2K activity has been shown in some cancers, such as breast cancer (Parmer *et al.* 1999). In contrast, overexpression of eEF2 protein was detected in over 90% of gastric and colorectal cancers (Nakamura *et al.* 2009). Moreover, overexpression of eEF2 in a gastric cancer cell line significantly enhanced cell growth and *in vivo* tumorigenicity in a mouse xenograft model (Nakamura *et al.* 2009). Hypothesizing that the identified point mutation could change the enzymatic activity of eEF2K, we recently started to test the kinase activity both in prokaryotic and in eukaryotic cells.

Furthermore, we would also like to test the effect of a mutation in the RFC1 gene (replication factor C 1). As the mutation in RFC1 is located in the PCNA binding domain, a first test would be to look at the capacity of wild type and mutated RFC1 to load PCNA and polymerase efficiently (Figure 6.1.B).

Of course, the above mutations only represent three examples out of a list of more than 400 mutations. The present challenge is to test the functional consequences of more of these mutations, in order to identify the driver mutation(s) in each PCa patient.

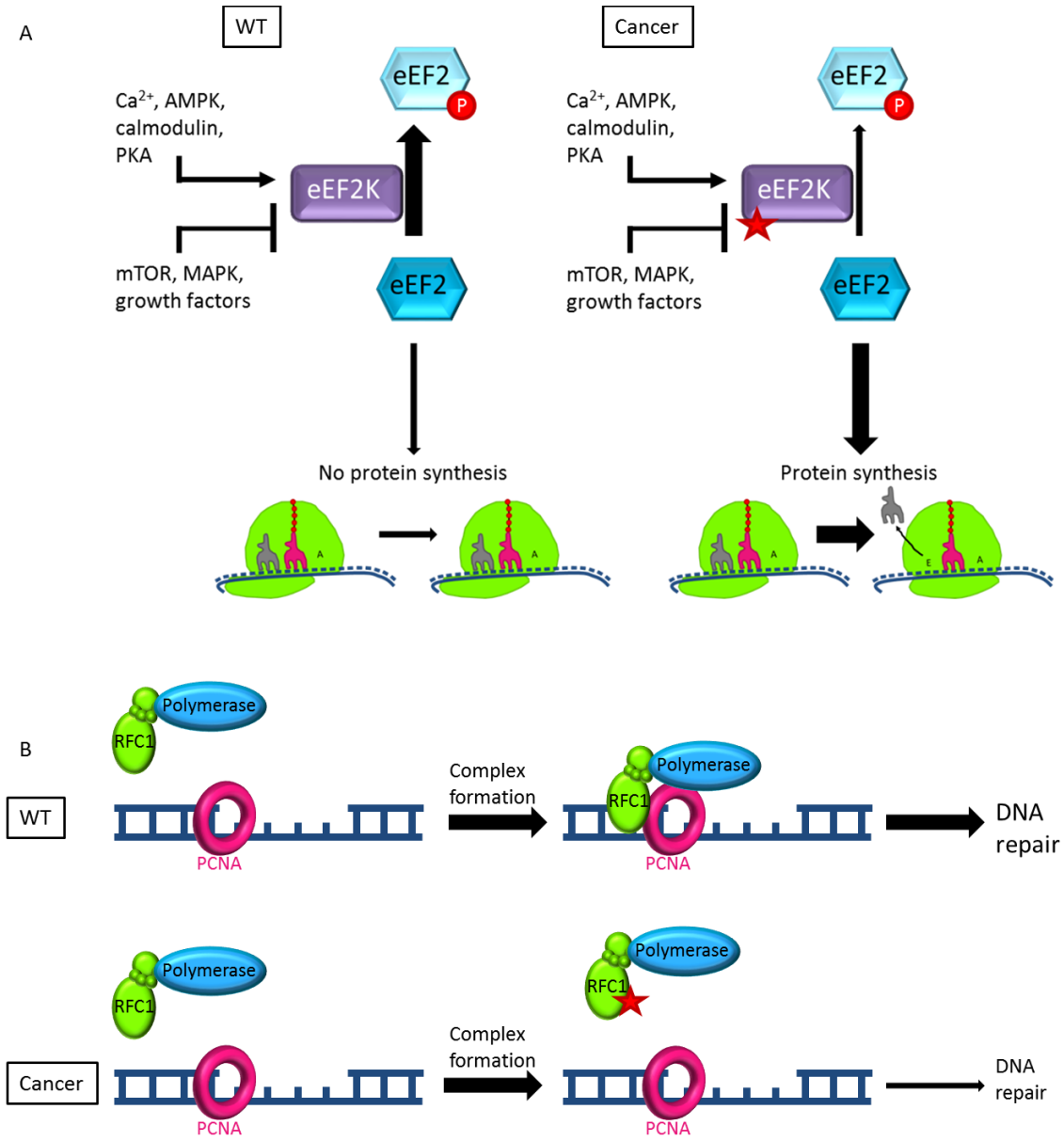



**Figure 6.1.** Hypothetical mechanisms of action of point mutations detected in high-risk PCa samples. **A.** In non-tumoral cells, eEF2 is phosphorylated and hence inactivated by eEF2K. As unphosphorylated eEF2 is required for the translation elongation, protein synthesis in these cells is limited. The detected mutation in eEF2K could inactivate the kinase, resulting in more active eEF2 and enhanced protein synthesis. **B.** In normal circumstances, RFC1 is required for the recruitment of DNA polymerase to PCNA (proliferating cell nuclear antigen) to start DNA repair. The mutation in RFC1 could result in an impaired recruitment of the polymerase to the site of DNA repair.

### *6.2.3. Goals of the PEARL consortium*

The high-risk PCa samples that were used in this study were collected within the frame of the PEARL (Prostate Cancer Research Leuven) consortium. The aim of the PEARL consortium is to prospectively integrate available data on exome sequencing, copy number alterations, lipidomics and DNA

methylation for a homogeneous cohort of high-risk PCa patients. The integration of these data with the data from the clinical research groups should allow the identification of biomarkers for PCa aggressiveness or lethality.

Combining our patient exomes with the already published exomes, results in over 200 PCa samples. Despite the heterogeneity of PCa, this already gives us a view on the repertoire of protein-coding mutations. Nevertheless, the listing of all mutations is only the beginning of new research, as a lot of questions still need to be answered. Which mutations are drivers, and which are passengers? How does each driver mutation exerts its effect? Is it through changes in gene expression or changes in enzymatic activity? Which cellular processes will be affected and will this result in a proliferative advantage for the cancer cells? If this is true, how will they do this? Will certain mutation profiles or affected pathways result in the identification of different PCa subtypes that have different clinical properties?

## 6.3. MASSIVELY PARALLEL SEQUENCING

### *6.3.1. What is next with massively parallel sequencing?*

Nowadays, most oncogenic variants that are identified are located in the coding genome, while the variation in the non-coding genome is largely ignored. The reason for this biased focus is purely practical: the effect of a coding mutation can be estimated or tested, while it is much more difficult, if not impossible, to determine the effect of a non-coding variant. Recently however, progress has been made to improve our understanding of the non-coding genome: The ENCODE project has assigned a biochemical function to 80% of the genome (Consortium *et al.* 2012). Incorporation of this knowledge will shift the discovery scale to the whole genome.

Oncogenesis is a complicated integration of alterations of multiple signal transduction pathways. They are the result of changes at different levels including the genome, epigenome, transcriptome, proteome, metabolome, lipidome and changes in chromatin and histones. It remains to be determined to what extent mutations in microRNAs and *cis*-regulatory elements (for example a promoter region with a mutation in a transcription factor binding site) contribute to the oncogenic process (Aerts *et al.* 2013). An example here is a recurrent mutation in the core promoter of the TERT gene as detected in melanoma, that generates a binding site for ETS transcription factors which then leads to increased transcriptional activity of the TERT promoter (Huang *et al.* 2013).

We hope that the integration of different data sets for genome, epigenome and transcriptome will help us to unravel the intricate regulatory connections between genetics and cancer. This will add an additional layer of knowledge, as is exemplified in the study of Sharma *et al.* where cancer-specific changes in DNA binding by the AR as well as other transcription factors were identified (Sharma *et al.* 2013).

### *6.3.2. Implementation of personalized genomics in prostate cancer*

Personalized medicine is defined as the individual treatment of patients, by targeting the pathways that are present or activated in their tumors. In PCa as in many other cancers, there is a considerable amount of interindividual tumor heterogeneity both at the genetic and epigenetic level. This heterogeneity challenges the 'one-fits-all' approaches for cancer management and highlights the need for individualized treatment approaches. It is therefore a future challenge to divide patients in subgroups and then look per subgroup to define what is treatable and what is curable. Such a personalized approach to cancer care is often superior to the current non-personalized standard therapies. Hence, we need to stratify patients prognostically, not predictively. This stratification will most likely be based on genomic features.

So far, molecular, clinical and pathologic research has focused on one tumor type at a time. However, due to the Cancer Genome Atlas Research Network (TCGA), it is now clear that cancers of disparate organs have many shared features, whereas, conversely, cancers from the same organ are often quite distinct (Cancer Genome Atlas Research *et al.* 2013). This implicates that somatic mutations that are rare in one tumor type, can be identified as drivers by aggregating events across different tumor types. This will improve detection of hotspot mutations which will ultimately lead to the identification of potential new drug targets. In other words, instead of treatments specific for PCa, it could become possible to treat one PCa in the same way as a colorectal cancer, if the same pathway in both patients is affected. This will only be possible by the study of larger cohorts of patients, as well as by a better understanding of the functional consequences of the identified alterations.

Ultimately, the mutations in each individual PCa case and of course also of the metastatic disease should be identified by genomic pathology and these data should be taken into consideration during the personalized treatment of the patient. However, a lot of progression is still needed: the price of sequencing has to drop further, the techniques to take liquid biopsies in the form of circulating tumor cells or DNA should be enhanced, and last but not least, new therapies have to be developed that can make use of the mutations that occur in the cancer. Efforts are made at the moment to inhibit PI3K (activated by PTEN mutation), inhibit mutant ARs resistant to enzalutamide and

abiraterone, development of synthetic lethal strategies for p53, … It will be exciting to see the continuing evolution in these different domains.

In conclusion, the introduction of next-generation sequencing in the clinic is an important step forward, even when much more work is still required to fully understand the oncogenome and to integrate it with the other available system biology approaches.

## SUMMARY

PCa is the second most frequently diagnosed cancer in males worldwide. A wide range of genomic alterations, including point mutations, copy number changes and rearrangements, can lead to the development of cancer. Due to the heterogeneity of PCa, it still remains a clinical challenge to differentiate indolent from very aggressive tumors. A better molecular profiling of the primary tumors should enable a better classification of the disease, ultimately providing information that could direct a more personalized treatment. One approach is to study the contribution of somatic base pair substitutions to the oncogenic process.

Cancer cell lines are commonly used as laboratory resources to study basic molecular and cellular biology. For prostate cancer (PCa), LNCaPs are the most commonly used cells. However, information on protein-changing mutations, genetic heterogeneity and genetic (in)stability is largely lacking for these cells. **In a first part**, we used exome-sequencing to focus on missense and nonsense single nucleotide variants and short insertions and deletions. We detected 1802 non-synonymous point mutations and 218 small insertions and deletions. While most detected mutations were undescribed so far, we confirmed the known mutations in the androgen receptor and the PTEN gene. Surprisingly, we confirmed 38 out of 42 mutations in DNA and RNA from different monoclonal and polyclonal LNCaP derivatives. From this, we deduced that LNCaP cells are heterozygous for a large number of variants and that both the variant and the wild type allele can be simultaneously expressed as mRNA. The fact that mutations in the E-cadherin, CDK4, Notch1 and PlexinB1 genes are absent in some subclones, strongly indicates a degree of genetic instability. Finally, to help identify the mutations that are most likely drivers of the oncogenic process, we developed an *in silico* protocol, which can be adopted for other exome analyses. We provided an extensive database of genetic variations in the exome of LNCaP cells, and these should be taken into consideration when using LNCaPs as a model for PCa.

The progression of PCa from androgen-dependent to androgen-independent poses an important clinical question, as the mechanisms leading to metastatic PCa are not well understood. C4-2B cells are derivatives of LNCaP cells, as they were derived from a bone metastasis that grew in nude mice after inoculation with the LNCaP-derived, castration-resistant C4-2 cells. The combination of LNCaP and C4-2B cells thus forms an excellent preclinical model to study the development of metastatic castration-resistant PCa. Because of the importance of this progression model, **a second part of this study** characterized both cell lines more thoroughly using exome and transcriptome sequencing to obtain point mutations and differential expression patterns. Exome sequencing detected 2188 and

3840 mutations in LNCaP and C4-2B cells respectively, of which 1784 were found in both cell lines. The use of more recent algorithms resulted in a higher sensitivity to detect point mutations, increasing the number of mutations detected in LNCaP from 1802 to 2188. Surprisingly, the parental LNCaP cells contained more than 400 mutations that were not found in the C4-2B exome. Moreover, more than half of the mutations found in the exomes of both cell lines were confirmed by analyzing the transcriptome sequencing data. The transcriptome data also revealed that 457 genes show increased expression and 246 genes show decreased expression in C4-2B cells as compared to LNCaP cells. Based on the list of C4-2B-specific point mutations and the list of differentially expressed genes, we detected changes in the focal adhesion and ECM-receptor interaction pathways which converged on the myosin light chain kinase gene. Whether this contributes to the metastatic potential of C4-2B cells remains to be investigated. To conclude, we provide lists of mutated genes and differentially expressed genes in the LNCaP and C4-2B PCa cell lines to all researchers interested in using these cells as preclinical models.

**A final component of this project** used tissue from 27 patients with high-risk primary PCa. We performed exome sequencing and copy number profiling of 27 primary prostate tumors and their normal tissue pairs. Tumors having a PSA > 20 ng/ml, or Gleason score ≥ 8 or clinical stage ≥ T2c are known to have a high risk on disease recurrence after treatment. In addition to amplifications and deletions that were described before, we identified a novel recurrent amplification on 7p22.3. Exome sequencing revealed one hypermutated sample containing 451 mutations, compared to an average of 19 mutations in the other samples, indicating that DNA repair is compromised in this sample. This hypermutated tumor indeed harbored a mutation in the DNA-repair gene Replication Factor C. This mutation is predicted to affect the interaction with PCNA and hence the recruitment of DNA polymerase to PCNA. In a second tumor, we detected a novel point mutation in the TET1 gene. This methylcytosine dioxygenase converts 5-methylcytosine to 5-hydroxymethylcytosine, which leads to demethylation of cytosines and might lead to changes in gene expression or chromatin organization. Immunoprecipitation of methylated and hydroxymethylated DNA followed by deep-sequencing performed on the tumor sample containing the A1908S TET1 mutation demonstrated an overall hypo-hydroxymethylation and hypermethylation at specific genomic loci, when compared to two other tumor samples without mutation in TET1. This effect was corroborated by the *in vitro* effect of the mutation on the dioxygenase activity as assessed by dot blot assays. We further showed that there is an overlap between groups of androgen-regulated genes and the TET1-regulated genes. Moreover, in cotransfection experiments, TET1 seems to act as a coactivator of the androgen receptor. From the above data, we conclude that the A1908S TET1 mutation as detected in a primary PCa leads to partial loss of TET1 tumor suppressor activity.

## SAMENVATTING

PCa is, wereldwijd, de op één na meest frequent gediagnosticeerde kanker in mannen. Een groot aantal genomische veranderingen, zoals puntmutaties, veranderingen in aantal kopijen van een (stuk van een) chromosoom of herschikkingen van stukken van chromosomen, kunnen leiden tot het ontstaan van kanker. Door de heterogeniteit van PCa, is het nog steeds een uitdaging in de kliniek om een onderscheid te kunnen maken tussen indolente en agressieve tumoren. Een betere kennis van de moleculaire processen in deze tumoren kan uiteindelijk leiden tot een betere classificatie van de ziekte, en dit kan uiteindelijk resulteren in een meer gepersonaliseerde behandeling. Een eerste aanzet hiertoe is om de bijdrage van puntmutaties te onderzoeken.

Kankercellijnen worden vaak gebruikt in het laboratorium om basisonderzoek te voeren naar moleculaire en cellulaire biologie. Voor prostaatkanker (PCa) zijn LNCaP cellen de meest gebruikte cellen. Toch is er nog niet veel geweten over mutaties, genetische heterogeniteit en instabiliteit in deze cellijn. Met behulp van exoom sequenties concentreerden we ons **in een eerste onderdeel** op puntmutaties die een aminozuur veranderen of die een stopcodon introduceren in een open leesraam of op korte inserties en deleties. Op deze manier detecteerden we 1802 puntmutaties en 218 korte inserties en deleties. De meerderheid van de gedetecteerde mutaties waren tot nog toe onbekend, hoewel we ook de reeds gekende mutaties in de androgeenreceptor en in het PTEN gen terugvonden. Daarenboven konden we 38 van de 42 geteste mutaties bevestigen in zowel monoklonale als polyklonale afgeleiden van LNCaP. Dit leidde tot de conclusie dat LNCaP cellen heterozygoot zijn voor een groot aantal van de gevonden mutaties en dat zowel het mutante als het wild type allel tegelijkertijd tot expressie kunnen komen. Het feit dat we de mutaties in E-cadherine, CDK4, Notch1 en PlexinB1 niet terugvinden in bepaalde LNCaP subklonen, wijst er sterk op dat de LNCaPs genetisch instabiel zijn. Tot slot hebben we een *in silico* protocol ontwikkeld dat kan helpen bij het identificeren van mutaties die waarschijnlijk bijdragen tot het oncogeen proces. Dit protocol kan ook gebruikt worden bij andere exoom analyses. We stellen een uitgebreide databank van genetische variaties ter beschikking die we terugvonden in het exoom van de LNCaP cellen. Met de mutaties in deze lijst wordt best rekening gehouden als men LNCaP cellen gebruikt als model voor PCa.

De mechanismen achter de evolutie van androgeen-afhankelijke naar androgeen-onafhankelijke PCa vormt een belangrijke klinische vraag, omdat de mechanismen die leiden tot metastatische PCa tot op heden niet goed begrepen zijn. C4-2B cellen zijn afgeleiden van de LNCaP cellen, aangezien ze werden geïsoleerd uit een botmetastase die groeide in naakte muizen nadat deze geïnoculeerd

werden met de castratie-resistente, van LNCaP cellen afgeleide, C4-2 cellen. De combinatie van LNCaP en C4-2B cellen vormt dus een uitstekend preklinisch model om de ontwikkeling van metastatische, castratie-resistente PCa te bestuderen. **In een tweede project** hebben we beide cellijnen uitgebreid gekarakteriseerd op het niveau van hun exoom en transcriptoom. Dit leidde tot de detectie van 2188 en 3840 puntmutaties in LNCaP en C4-2B cellen respectievelijk. Hiervan werden 1784 mutaties gevonden in beide cellijnen. Doordat er bij deze karakterisatie recentere algoritmen gebruikt werden dan bij de eerdere analyse van LNCaP cellen, detecteerden we een hoger aantal mutaties: 2188 in plaats van 1802. Tot onze verbazing detecteerden we in de LNCaP cellen meer dan 400 mutaties die niet werden teruggevonden in het exoom van C4-2B cellen. Door de sequenties van het exoom en het transcriptoom te vergelijken, konden we meer dan de helft van de mutaties die we detecteerden in het exoom ook terugvinden in het transcriptoom, en dit voor beide cellijnen. Daarenboven konden we uit het transcriptoom besluiten dat 457 genen verhoogd en 246 genen minder tot expressie kwamen in C4-2B cellen wanneer deze vergeleken werden met de expressie in LNCaP cellen. Door gebruik te maken van C4-2B-specifieke puntmutaties en dit te combineren met de lijst van genen die differentieel tot expressie komen, probeerden we een aantal verschillen terug te vinden die de grotere metastatische eigenschappen van C4-2B cellen zouden kunnen verklaren. We detecteerden veranderingen in adhesie en in de ECM-receptor interacties. Deze signaalcascades oefenen beiden invloed uit op het MLCK gen (myosin light chain kinase) wat de metastatische eigenschappen van C4-2B cellen zou kunnen verklaren. Deze studie resulteerde uiteindelijk in lijsten van gemuteerde genen en genen die differentieel tot expressie komen in de LNCaP en in de C4-2B cellijn.

**In een derde onderdeel** van deze studie beschikten we over weefsels van 27 patiënten die in de kliniek momenteel worden ingeschat als hebbende een ‘hoog-risico’ op het terugkeren van de ziekte. Van deze 27 primaire tumore werd het exoom in sequentie gezet en amplificaties en deleties bepaald. We detecteerden verschillende recurrente genoomamplificaties en deleties die reeds gekend waren, maar we detecteerden ook een ongekende recurrente amplificatie op 7p22.3. De exoomsequenties toonden aan dat er één staal gehypermuteerd was: het bevatte 451 puntmutaties, terwijl er in de andere stalen gemiddeld slechts 19 mutaties teruggevonden werden. Dit zou er kunnen op wijzen dat het DNA herstel beïnvloed is in deze tumor. Inderdaad, we vonden een mutatie in het RFC1 gen (Replication Factor C 1) dat betrokken is bij het herstel van DNA. Van de mutatie in dit gen wordt voorspeld dat ze de interactie met PCNA zal beïnvloeden en dit zal op zijn beurt de interactie met het DNA polymerase beïnvloeden. In een tweede tumor vonden we een mutatie in het TET1 gen. Dit gen codeert voor een methylcytosine dioxygenase dat 5-methylcytosine omvormt tot 5-hydroxymethylcytosine, wat uiteindelijk resulteert in demethylatie en veranderingen in

genexpressie. Een eerste experiment bestond uit de immunoprecipitatie van gemethyleerd en hydroxygemethyleerd DNA gevolgd door het in sequentie zetten van dit DNA. Het tumorstaal met de A1908S mutatie in TET1 toonde in het algemeen een hypo-hydroxymethylatie en een hypermethylatie op bepaalde genomische posities, wanneer dit vergeleken werd met twee tumorstalen zonder mutatie in TET1. Deze resultaten werden nog versterkt bij de dot blot experimenten, die het effect van de mutatie op de dioxygenase activiteit *in vitro* nakijken. TET1 lijkt ook een coactivator van de androgeenreceptor te zijn. Uit bovenstaande data concluderen we dat de A1908S mutatie in TET1 resulteert in een gedeeltelijk verlies van de TET1 tumor suppressor activiteit.

## REFERENCES


Adzhubei, I. A., Schmidt, S., Peshkin, L.*, et al.* (2010). "A method and server for predicting damaging missense mutations." Nat methods **7**(4): 248-249.

Aerts, S. and Cools, J. (2013). "Cancer: Mutations close in on gene regulation." Nature **499**(7456): 35-36.

Ahmadiyeh, N., Pomerantz, M. M., Grisanzio, C.*, et al.* (2010). "8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC." Proc Natl Acad Sci U S A **107**(21): 9742-9746.

Albers, C. A., Lunter, G., MacArthur, D. G.*, et al.* (2011). "Dindel: accurate indel calls from short-read data." Genome res **21**(6): 961-973.

Amin Al Olama, A., Kote-Jarai, Z., Schumacher, F. R.*, et al.* (2013). "A meta-analysis of genome-wide association studies to identify prostate cancer susceptibility loci associated with aggressive and non-aggressive disease." Hum Mol Genet **22**(2): 408-415.

An, J., Wang, C., Deng, Y.*, et al.* (2014). "Destruction of Full-Length Androgen Receptor by Wild-Type SPOP, but Not Prostate-Cancer-Associated Mutants." Cell Rep **6**(4): 657-669.

Aryee, M. J., Liu, W., Engelmann, J. C.*, et al.* (2013). "DNA methylation alterations exhibit intraindividual stability and interindividual heterogeneity in prostate cancer metastases." Sci Transl Med **5**(169): 169ra110.

Assinder, S. J., Dong, Q., Kovacevic, Z.*, et al.* (2009). "The TGF-beta, PI3K/Akt and PTEN pathways: established and proposed biochemical integration in prostate cancer." Biochem J **417**(2): 411-421.

Attard, G., Clark, J., Ambroisine, L.*, et al.* (2008). "Heterogeneity and clinical significance of ETV1 translocations in human prostate cancer." Br J Cancer **99**(2): 314-320.

Baca, S. C., Prandi, D., Lawrence, M. S.*, et al.* (2013). "Punctuated evolution of prostate cancer genomes." Cell **153**(3): 666-677.

Bainbridge, M. N., Warren, R. L., Hirst, M.*, et al.* (2006). "Analysis of the prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach." BMC genomics **7**: 246.

Balk, S. P. and Knudsen, K. E. (2008). "AR, the cell cycle, and prostate cancer." Nucl Recept Signal **6**: e001.

Barbieri, C. E., Baca, S. C., Lawrence, M. S.*, et al.* (2012). "Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer." Nat Genet **44**(6): 685-689.

Barbieri, C. E., Bangma, C. H., Bjartell, A.*, et al.* (2013). "The mutational landscape of prostate cancer." Eur Urol **64**(4): 567-576.

Bardelli, A., Parsons, D. W., Silliman, N.*, et al.* (2003). "Mutational analysis of the tyrosine kinome in colorectal cancers." Science **300**(5621): 949.

Baylin, S. B. and Jones, P. A. (2011). "A decade of exploring the cancer epigenome - biological and translational implications." Nat Rev Cancer **11**(10): 726-734.

Beltran, H., Yelensky, R., Frampton, G. M.*, et al.* (2012). "Targeted Next-generation Sequencing of Advanced Prostate Cancer Identifies Potential Therapeutic Targets and Disease Heterogeneity." Eur Urol **63**(5): 920-926.

Berger, A. H., Knudson, A. G. and Pandolfi, P. P. (2011a). "A continuum model for tumour suppression." Nature **476**(7359): 163-169.

Berger, M. F., Lawrence, M. S., Demichelis, F.*, et al.* (2011b). "The genomic complexity of primary human prostate cancer." Nature **470**(7333): 214-220.

Beroukhim, R., Getz, G., Nghiemphu, L.*, et al.* (2007). "Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma." Proc Natl Acad Sci U S A **104**(50): 20007-20012.

Bisoffi, M., Klima, I., Gresko, E.*, et al.* (2004). "Expression profiles of androgen independent bone metastatic prostate cancer cells indicate up-regulation of the putative serine-threonine kinase GS3955." J Urol **172**(3): 1145-1150.

Blankenberg, D., Von Kuster, G., Coraor, N.*, et al.* (2010). "Galaxy: a web-based genome analysis tool for experimentalists." Curr Protoc Mol Biol **Chapter 19**: Unit 19 10 11-21.

Blazek, D., Kohoutek, J., Bartholomeeusen, K.*, et al.* (2011). "The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes." Genes Dev **25**(20): 2158-2172.

Borno, S. T., Fischer, A., Kerick, M.*, et al.* (2012). "Genome-wide DNA methylation events in TMPRSS2-ERG fusion-negative prostate cancers implicate an EZH2-dependent mechanism with miR-26a hypermethylation." Cancer Discov **2**(11): 1024-1035.

Boyd, L. K., Mao, X., Xue, L.*, et al.* (2012). "High-resolution genome-wide copy-number analysis suggests a monoclonal origin of multifocal prostate cancer." Genes Chromosomes Cancer **51**(6): 579-589.

Bratt, O., Damber, J. E., Emanuelsson, M.*, et al.* (2002). "Hereditary prostate cancer: clinical characteristics and survival." J Urol **167**(6): 2423-2426.

Briganti, A., Joniau, S., Gontero, P.*, et al.* (2012). "Identifying the best candidate for radical prostatectomy among patients with high-risk prostate cancer." Eur Urol **61**(3): 584-592.

Bussemakers, M. J., van Bokhoven, A., Verhaegh, G. W.*, et al.* (1999). "DD3: a new prostate-specific gene, highly overexpressed in prostate cancer." Cancer Res **59**(23): 5975-5979.

Cadet, J. and Wagner, J. R. (2014). "TET enzymatic oxidation of 5-methylcytosine, 5-hydroxymethylcytosine and 5-formylcytosine." Mutat Res **764-765**: 18-35.

Cancer Genome Atlas Research, N., Weinstein, J. N., Collisson, E. A.*, et al.* (2013). "The Cancer Genome Atlas Pan-Cancer analysis project." Nat Genet **45**(10): 1113-1120.

Carlsson, J., Davidsson, S., Helenius, G.*, et al.* (2011). "A miRNA expression signature that separates between normal and malignant prostate tissues." Cancer Cell Int **11**(1): 14.

Carver, B. S., Tran, J., Gopalan, A.*, et al.* (2009). "Aberrant ERG expression cooperates with loss of PTEN to promote cancer progression in the prostate." Nat Genet **41**(5): 619-624.

Casey, O. M., Fang, L., Hynes, P. G.*, et al.* (2012). "TMPRSS2- driven ERG expression in vivo increases self-renewal and maintains expression in a castration resistant subpopulation." PLoS One **7**(7): e41668.

Centenera, M. M., Gillis, J. L., Hanson, A. R.*, et al.* (2012). "Evidence for efficacy of new Hsp90 inhibitors revealed by ex vivo culture of human prostate tumors." Clin Cancer Res **18**(13): 3562-3570.

Centenera, M. M., Raj, G. V., Knudsen, K. E.*, et al.* (2013). "Ex vivo culture of human prostate tissue and drug development." Nat Rev Urol **10**(8): 483-487.

Chang, H., Jackson, D. G., Kayne, P. S.*, et al.* (2011). "Exome Sequencing Reveals Comprehensive Genomic Alterations across Eight Cancer Cell Lines." PloS one **6**(6): e21097.

Chen, Y., Wang, J., Fraig, M. M.*, et al.* (2001). "Defects of DNA mismatch repair in human prostate cancer." Cancer Res **61**(10): 4112-4121.

Cheng, I., Levin, A. M., Tai, Y. C.*, et al.* (2012). "Copy number alterations in prostate tumors and disease aggressiveness." Genes Chromosomes Cancer **51**(1): 66-76.

Choudhury, A. D., Eeles, R., Freedland, S. J.*, et al.* (2012). "The role of genetic markers in the management of prostate cancer." Eur Urol **62**(4): 577-587.

Chung, S., Nakagawa, H., Uemura, M.*, et al.* (2011). "Association of a novel long non-coding RNA in 8q24 with prostate cancer susceptibility." Cancer Sci **102**(1): 245-252.

Cibulskis, K., Lawrence, M. S., Carter, S. L.*, et al.* (2013). "Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples." Nat Biotechnol **31**(3): 213-219.

Clark, M. J., Homer, N., O'Connor, B. D.*, et al.* (2010). "U87MG decoded: the genomic sequence of a cytogenetically aberrant human cancer cell line." PLoS Genet **6**(1): e1000832.

Clinckemalie, L., Spans, L., Dubois, V.*, et al.* (2013). "Androgen regulation of the TMPRSS2 gene and the effect of a SNP in an androgen response element." Mol Endocrinol **27**(12): 2028-2040.

Consortium, E. P., Bernstein, B. E., Birney, E.*, et al.* (2012). "An integrated encyclopedia of DNA elements in the human genome." Nature **489**(7414): 57-74.

Cooperberg, M. R., Moul, J. W. and Carroll, P. R. (2005). "The changing face of prostate cancer." J Clin Oncol **23**(32): 8146-8151.

Cramer, S. D., Chang, B. L., Rao, A.*, et al.* (2003). "Association between genetic polymorphisms in the prostate-specific antigen gene promoter and serum prostate-specific antigen levels." J Natl Cancer Inst **95**(14): 1044-1053.

Cui, W. J., Liu, Y., Zhou, X. L.*, et al.* (2010). "Myosin light chain kinase is responsible for high proliferative ability of breast cancer cells via anti-apoptosis involving p38 pathway." Acta Pharmacol Sin **31**(6): 725-732.

Cui, Z., Ren, S., Lu, J.*, et al.* (2013). "The prostate cancer-up-regulated long noncoding RNA PlncRNA-1 modulates apoptosis and proliferation through reciprocal regulation of androgen receptor." Urol Oncol **31**(7): 1117-1123.

D'Amico, A. V., Whittington, R., Malkowicz, S. B.*, et al.* (1998). "Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer." JAMA **280**(11): 969-974.

Demichelis, F., Setlur, S. R., Beroukhim, R.*, et al.* (2009). "Distinct genomic aberrations associated with ERG rearranged prostate cancer." Genes Chromosomes Cancer **48**(4): 366-380.

Denayer, S., Helsen, C., Thorrez, L.*, et al.* (2010). "The rules of DNA recognition by the androgen receptor." Mol Endocrinol **24**(5): 898-913.

Dominissini, D., Moshitch-Moshkovitz, S., Amariglio, N.*, et al.* (2011). "Adenosine-to-inosine RNA editing meets cancer." Carcinogenesis **32**(11): 1569-1577.

Donaldson, L. W., Petersen, J. M., Graves, B. J.*, et al.* (1994). "Secondary structure of the ETS domain places murine Ets-1 in the superfamily of winged helix-turn-helix DNA-binding proteins." Biochemistry **33**(46): 13509-13516.

Dong, J. T. (2006). "Prevalent mutations in prostate cancer." J Cell Biochem **97**(3): 433-447.

Draghici, S., Khatri, P., Tarca, A. L.*, et al.* (2007). "A systems biology approach for pathway level analysis." Genome Res **17**(10): 1537-1545.

Duggan, D., Zheng, S. L., Knowlton, M.*, et al.* (2007). "Two genome-wide association studies of aggressive prostate cancer implicate putative prostate tumor suppressor gene DAB2IP." J Natl Cancer Inst **99**(24): 1836-1844.

Eeles, R. A., Olama, A. A., Benlloch, S.*, et al.* (2013). "Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array." Nat Genet **45**(4): 385-391.

Eifler, J. B., Feng, Z., Lin, B. M.*, et al.* (2013). "An updated prostate cancer staging nomogram (Partin tables) based on cases from 2006 to 2011." BJU Int **111**(1): 22-29.

Epstein, J. I., Allsbrook, W. C., Jr., Amin, M. B.*, et al.* (2005). "The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma." Am J Surg Pathol **29**(9): 1228-1242.

Esgueva, R., Perner, S., C, J. L.*, et al.* (2010). "Prevalence of TMPRSS2-ERG and SLC45A3-ERG gene fusions in a large prostatectomy cohort." Mod Pathol **23**(4): 539-546.

Esquenet, M., Swinnen, J. V., Heyns, W.*, et al.* (1997). "LNCaP prostatic adenocarcinoma cells derived from low and high passage numbers display divergent responses not only to androgens but also to retinoids." J Steroid Biochem Mol Biol **62**(5-6): 391-399.

Fazal, F., Gu, L., Ihnatovych, I.*, et al.* (2005). "Inhibiting myosin light chain kinase induces apoptosis in vitro and in vivo." Mol Cell Biol **25**(14): 6259-6266.

Ferlay, J., Shin, H. R., Bray, F.*, et al.* (2010). "Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008." Int J Cancer **127**(12): 2893-2917.

Ferlay, J., Steliarova-Foucher, E., Lortet-Tieulent, J.*, et al.* (2013). "Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012." Eur J Cancer **49**(6): 1374-1403.

Figueroa, M. E., Abdel-Wahab, O., Lu, C.*, et al.* (2010). "Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation." Cancer Cell **18**(6): 553-567.

FitzGerald, L. M., Kwon, E. M., Conomos, M. P.*, et al.* (2011). "Genome-wide association study identifies a genetic variant associated with risk for more aggressive prostate cancer." Cancer Epidemiol Biomarkers Prev **20**(6): 1196-1203.

Fu, Z., Dozmorov, I. M. and Keller, E. T. (2002). "Osteoblasts produce soluble factors that induce a gene expression pattern in non-metastatic prostate cancer cells, similar to that found in bone metastatic prostate cancer cells." Prostate **51**(1): 10-20.

Fujita, A., Gomes, L. R., Sato, J. R.*, et al.* (2008). "Multivariate gene expression analysis reveals functional connectivity changes between normal/tumoral prostates." BMC Syst Biol **2**: 106.

Fuse, M., Kojima, S., Enokida, H.*, et al.* (2012). "Tumor suppressive microRNAs (miR-222 and miR-31) regulate molecular pathways based on microRNA expression signature in prostate cancer." J Hum Genet **57**(11): 691-699.

Goecks, J., Nekrutenko, A. and Taylor, J. (2010). "Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences." Genome Biol **11**(8): R86.

Goh, C. L., Schumacher, F. R., Easton, D.*, et al.* (2012). "Genetic variants associated with predisposition to prostate cancer and potential clinical implications." J Intern Med **271**(4): 353-365.

Gordon, V., Bhadel, S., Wunderlich, W.*, et al.* (2010). "CDK9 regulates AR promoter selectivity and cell growth through serine 81 phosphorylation." Mol Endocrinol **24**(12): 2267-2280.

Grasso, C. S., Wu, Y. M., Robinson, D. R.*, et al.* (2012). "The mutational landscape of lethal castration-resistant prostate cancer." Nature **487**(7406): 239-243.

Greenman, C., Stephens, P., Smith, R.*, et al.* (2007). "Patterns of somatic mutation in human cancer genomes." Nature **446**(7132): 153-158.

Gregory, P. A., Bert, A. G., Paterson, E. L.*, et al.* (2008). "The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1." Nat Cell Biol **10**(5): 593-601.

Grindedal, E. M., Moller, P., Eeles, R.*, et al.* (2009). "Germ-line mutations in mismatch repair genes associated with prostate cancer." Cancer Epidemiol Biomarkers Prev **18**(9): 2460-2467.

Guo, J. U., Su, Y., Zhong, C.*, et al.* (2011). "Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain." Cell **145**(3): 423-434.

Haelens, A., Verrijdt, G., Callewaert, L.*, et al.* (2001). "Androgen-receptor-specific DNA binding to an element in the first exon of the human secretory component gene." Biochem J **353**(Pt 3): 611-620.

Haffner, M. C., Chaux, A., Meeker, A. K.*, et al.* (2011). "Global 5-hydroxymethylcytosine content is significantly reduced in tissue stem/progenitor cell compartments and in human cancers." Oncotarget **2**(8): 627-637.

Haffner, M. C., Mosbruger, T., Esopi, D. M.*, et al.* (2013). "Tracking the clonal origin of lethal prostate cancer." J Clin Invest **123**(11): 4918-4922.

Han, B., Mehra, R., Dhanasekaran, S. M.*, et al.* (2008). "A fluorescence in situ hybridization screen for E26 transformation-specific aberrations: identification of DDX5-ETV4 fusion protein in prostate cancer." Cancer Res **68**(18): 7629-7637.

Hashiguchi, K., Matsumoto, Y. and Yasui, A. (2007). "Recruitment of DNA repair synthesis machinery to sites of DNA damage/repair in living human cells." Nucleic Acids Res **35**(9): 2913-2923.

He, Y. F., Li, B. Z., Li, Z.*, et al.* (2011). "Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA." Science **333**(6047): 1303-1307.

Heidenreich, A., Bastian, P. J., Bellmunt, J.*, et al.* (2013). "EAU Guidelines on Prostate Cancer. Part II: Treatment of Advanced, Relapsing, and Castration-Resistant Prostate Cancer." Eur Urol **65**(2): 467-479.

Heidenreich, A., Bastian, P. J., Bellmunt, J.*, et al.* (2014). "EAU Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent-Update 2013." Eur Urol **65**(1): 124-137.

Helgeson, B. E., Tomlins, S. A., Shah, N.*, et al.* (2008). "Characterization of TMPRSS2:ETV5 and SLC45A3:ETV5 gene fusions in prostate cancer." Cancer Res **68**(1): 73-80.

Hermans, K. G., Bressers, A. A., van der Korput, H. A.*, et al.* (2008a). "Two unique novel prostate-specific and androgen-regulated fusion partners of ETV4 in prostate cancer." Cancer Res **68**(9): 3094-3098.

Hermans, K. G., van der Korput, H. A., van Marion, R.*, et al.* (2008b). "Truncated ETV1, fused to novel tissue-specific genes, and full-length ETV1 in prostate cancer." Cancer Res **68**(18): 7541-7549.

Hizli, A. A., Chi, Y., Swanger, J.*, et al.* (2013). "Phosphorylation of eukaryotic elongation factor 2 (eEF2) by cyclin A-cyclin-dependent kinase 2 regulates its inhibition by eEF2 kinase." Mol Cell Biol **33**(3): 596-604.

Horoszewicz, J. S., Leong, S. S., Chu, T. M.*, et al.* (1980). "The LNCaP cell line--a new model for studies on human prostatic carcinoma." Prog Clin Biol Res **37**: 115-132.

Horoszewicz, J. S., Leong, S. S., Kawinski, E.*, et al.* (1983). "LNCaP model of human prostatic carcinoma." Cancer Res **43**(4): 1809-1818.

Hotchkiss, R. D. (1948). "The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography." J Biol Chem **175**(1): 315-332.

Howlader, N., Noone, A. M., Krapcho, M.*, et al.* (2013, April 2013). "SEER Cancer Statistics Review, 1975-2010." from http://seer.cancer.gov/csr/1975_2010/.

Hsu, C. H., Peng, K. L., Kang, M. L.*, et al.* (2012). "TET1 suppresses cancer invasion by activating the tissue inhibitors of metalloproteinases." Cell Rep **2**(3): 568-579.

Hsu, C. Y., Joniau, S., Oyen, R.*, et al.* (2007). "Outcome of surgery for clinical unilateral T3a prostate cancer: a single-institution experience." Eur Urol **51**(1): 121-128; discussion 128-129.

Huang da, W., Sherman, B. T. and Lempicki, R. A. (2009a). "Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists." Nucleic Acids Res **37**(1): 1-13.

Huang da, W., Sherman, B. T. and Lempicki, R. A. (2009b). "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources." Nat Protoc **4**(1): 44-57.

Huang, F. W., Hodis, E., Xu, M. J.*, et al.* (2013). "Highly recurrent TERT promoter mutations in human melanoma." Science **339**(6122): 957-959.

Hughes, C., Murphy, A., Martin, C.*, et al.* (2005). "Molecular pathology of prostate cancer." J Clin Pathol **58**(7): 673-684.

Ianculescu, I., Wu, D. Y., Siegmund, K. D.*, et al.* (2012). "Selective roles for cAMP response element-binding protein binding protein and p300 protein as coregulators for androgen-regulated gene expression in advanced prostate cancer cells." J Biol Chem **287**(6): 4000-4013.

Imamura, Y., Sakamoto, S., Endo, T.*, et al.* (2012). "FOXA1 promotes tumor progression in prostate cancer via the insulin-like growth factor binding protein 3 pathway." PLoS One **7**(8): e42456.

Ito, S., Shen, L., Dai, Q.*, et al.* (2011). "Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine." Science **333**(6047): 1300-1303.

Jin, G., Lu, L., Cooney, K. A.*, et al.* (2012). "Validation of prostate cancer risk-related loci identified from genome-wide association studies using family-based association analysis: evidence from the International Consortium for Prostate Cancer Genetics (ICPCG)." Hum Genet **131**(7): 1095-1103.

Jin, S. G., Jiang, Y., Qiu, R.*, et al.* (2011). "5-Hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations." Cancer Res **71**(24): 7360-7365.

Joniau, S., Van den Bergh, L., Lerut, E.*, et al.* (2013). "Mapping of pelvic lymph node metastases in prostate cancer." Eur Urol **63**(3): 450-458.

Kaighn, M. E., Narayan, K. S., Ohnuki, Y.*, et al.* (1979). "Establishment and characterization of a human prostatic carcinoma cell line (PC-3)." Invest Urol **17**(1): 16-23.

Kan, Z., Jaiswal, B. S., Stinson, J.*, et al.* (2010). "Diverse somatic mutation patterns and pathway alterations in human cancers." Nature **466**(7308): 869-873.

Kaneko, K., Satoh, K., Masamune, A.*, et al.* (2002). "Myosin light chain kinase inhibitors can block invasion and adhesion of human pancreatic cancer cell lines." Pancreas **24**(1): 34-41.

Kattan, M. W., Eastham, J. A., Stapleton, A. M.*, et al.* (1998). "A preoperative nomogram for disease recurrence following radical prostatectomy for prostate cancer." J Natl Cancer Inst **90**(10): 766-771.

Khuon, S., Liang, L., Dettman, R. W.*, et al.* (2010). "Myosin light chain kinase mediates transcellular intravasation of breast cancer cells through the underlying endothelial cells: a three-dimensional FRET study." J Cell Sci **123**(Pt 3): 431-440.

Kim, J. H., Dhanasekaran, S. M., Prensner, J. R.*, et al.* (2011a). "Deep sequencing reveals distinct patterns of DNA methylation in prostate cancer." Genome Res **21**(7): 1028-1041.

Kim, J. W., Kim, S. T., Turner, A. R.*, et al.* (2012). "Identification of new differentially methylated genes that have potential functional consequences in prostate cancer." PLoS One **7**(10): e48455.

Kim, S. J., Kelly, W. K., Fu, A.*, et al.* (2011b). "Genome-wide methylation analysis identifies involvement of TNF-alpha mediated cancer pathways in prostate cancer." Cancer Lett **302**(1): 47-53.

King, J. C., Xu, J., Wongvipat, J.*, et al.* (2009). "Cooperativity of TMPRSS2-ERG with PI3-kinase pathway activation in prostate oncogenesis." Nat Genet **41**(5): 524-526.

Klotz, L., Zhang, L., Lam, A.*, et al.* (2010). "Clinical results of long-term follow-up of a large, active surveillance cohort with localized prostate cancer." J Clin Oncol **28**(1): 126-131.

Kobayashi, Y., Absher, D. M., Gulzar, Z. G.*, et al.* (2011). "DNA methylation profiling reveals novel biomarkers and important roles for DNA methyltransferases in prostate cancer." Genome Res **21**(7): 1017-1027.

Koboldt, D. C., Chen, K., Wylie, T.*, et al.* (2009). "VarScan: variant detection in massively parallel sequencing of individual and pooled samples." Bioinformatics **25**(17): 2283-2285.

Korenchuk, S., Lehr, J. E., L, M. C.*, et al.* (2001). "VCaP, a cell-based model system of human prostate cancer." In Vivo **15**(2): 163-168.

Korpal, M., Lee, E. S., Hu, G.*, et al.* (2008). "The miR-200 family inhibits epithelial-mesenchymal transition and cancer cell migration by direct targeting of E-cadherin transcriptional repressors ZEB1 and ZEB2." J Biol Chem **283**(22): 14910-14914.

Kote-Jarai, Z., Saunders, E. J., Leongamornlert, D. A.*, et al.* (2013). "Fine-mapping identifies multiple prostate cancer risk loci at 5p15, one of which associates with TERT expression." Hum Mol Genet **22**(12): 2520-2528.

Kron, K., Pethe, V., Briollais, L.*, et al.* (2009). "Discovery of novel hypermethylated genes in prostate cancer using genomic CpG island microarrays." PLoS One **4**(3): e4830.

Krzywinski, M., Schein, J., Birol, I.*, et al.* (2009). "Circos: an information aesthetic for comparative genomics." Genome Res **19**(9): 1639-1645.

Kudo, Y., Tateishi, K., Yamamoto, K.*, et al.* (2012). "Loss of 5-hydroxymethylcytosine is accompanied with malignant cellular transformation." Cancer Sci **103**(4): 670-676.

Kumar, A., White, T. A., MacKenzie, A. P.*, et al.* (2011). "Exome sequencing identifies a spectrum of mutation frequencies in advanced and lethal prostate cancers." Proc Natl Acad Sci U S A **108**(41): 17087-17092.

Kumar, P., Henikoff, S. and Ng, P. C. (2009). "Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm." Nat Protoc **4**(7): 1073-1081.

Lang, G. I., Parsons, L. and Gammie, A. E. (2013). "Mutation rates, spectra, and genome-wide distribution of spontaneous mutations in mismatch repair deficient yeast." G3 (Bethesda) **3**(9): 1453-1465.

Langmead, B., Trapnell, C., Pop, M.*, et al.* (2009). "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome." Genome Biol **10**(3): R25.

Lapuk, A. V., Wu, C., Wyatt, A. W.*, et al.* (2012). "From sequence to molecular pathology, and a mechanism driving the neuroendocrine phenotype in prostate cancer." J Pathol **227**(3): 286-297.

Larson, D. E., Harris, C. C., Chen, K.*, et al.* (2012). "SomaticSniper: identification of somatic point mutations in whole genome sequencing data." Bioinformatics **28**(3): 311-317.

Lawrence, M. S., Stojanov, P., Mermel, C. H.*, et al.* (2014). "Discovery and saturation analysis of cancer genes across 21 tumour types." Nature **505**(7484): 495-501.

Leach, F. S., Velasco, A., Hsieh, J. T.*, et al.* (2000). "The mismatch repair gene hMSH2 is mutated in the prostate cancer cell line LNCaP." J Urol **164**(5): 1830-1833.

Lee, W. H., Morton, R. A., Epstein, J. I.*, et al.* (1994). "Cytidine methylation of regulatory sequences near the pi-class glutathione S-transferase gene accompanies human prostatic carcinogenesis." Proc Natl Acad Sci U S A **91**(24): 11733-11737.

Lee, Y. S., Kim, H. K., Chung, S.*, et al.* (2005). "Depletion of human micro-RNA miR-125b reveals that it is critical for the proliferation of differentiated cells but not for the down-regulation of putative targets during differentiation." J Biol Chem **280**(17): 16635-16641.

Li, H. and Durbin, R. (2009a). "Fast and accurate short read alignment with Burrows-Wheeler transform." Bioinformatics **25**(14): 1754-1760.

Li, H., Handsaker, B., Wysoker, A.*, et al.* (2009b). "The Sequence Alignment/Map format and SAMtools." Bioinformatics **25**(16): 2078-2079.

Li, T., Li, D., Sha, J.*, et al.* (2009c). "MicroRNA-21 directly targets MARCKS and promotes apoptosis resistance and invasion in prostate cancer cells." Biochem Biophys Res Commun **383**(3): 280-285.

Lichtenstein, P., Holm, N. V., Verkasalo, P. K.*, et al.* (2000). "Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland." N Engl J Med **343**(2): 78-85.

Lin, D., Wyatt, A. W., Xue, H.*, et al.* (2014). "High fidelity patient-derived xenografts for accelerating prostate cancer discovery and drug development." Cancer Res **74**(4): 1272-1283.

Lindberg, J., Klevebring, D., Liu, W.*, et al.* (2013). "Exome sequencing of prostate cancer supports the hypothesis of independent tumour origins." Eur Urol **63**(2): 347-353.

Lindberg, J., Mills, I. G., Klevebring, D.*, et al.* (2012). "The Mitochondrial and Autosomal Mutation Landscapes of Prostate Cancer." Eur Urol **63**(4): 702-708.

Liu, A. Y., Brubaker, K. D., Goo, Y. A.*, et al.* (2004). "Lineage relationship between LNCaP and LNCaP-derived prostate cancer cell lines." Prostate **60**(2): 98-108.

Liu, W., Laitinen, S., Khan, S.*, et al.* (2009). "Copy number analysis indicates monoclonal origin of lethal metastatic prostate cancer." Nat Med **15**(5): 559-565.

Liu, X., Han, S., Wang, Z.*, et al.* (2013). "Variant callers for next-generation sequencing data: a comparison study." PLoS One **8**(9): e75619.

Lonigro, R. J., Grasso, C. S., Robinson, D. R.*, et al.* (2011). "Detection of somatic copy number alterations in cancer using targeted exome capture sequencing." Neoplasia **13**(11): 1019-1025.

Lu-Yao, G. L., Albertsen, P. C., Moore, D. F.*, et al.* (2009). "Outcomes of localized prostate cancer following conservative management." JAMA **302**(11): 1202-1209.

MacInnis, R. J., Antoniou, A. C., Eeles, R. A.*, et al.* (2010). "Prostate cancer segregation analyses using 4390 families from UK and Australian population-based studies." Genet Epidemiol **34**(1): 42-50.

Maher, C. A., Kumar-Sinha, C., Cao, X.*, et al.* (2009a). "Transcriptome sequencing to detect gene fusions in cancer." Nature **458**(7234): 97-101.

Maher, C. A., Palanisamy, N., Brenner, J. C.*, et al.* (2009b). "Chimeric transcript discovery by paired-end transcriptome sequencing." Proc Natl Acad Sci U S A **106**(30): 12353-12358.

Majumdar, S., Buckles, E., Estrada, J.*, et al.* (2011). "Aberrant DNA methylation and prostate cancer." Curr Genomics **12**(7): 486-505.

Majumder, P. K., Grisanzio, C., O'Connell, F.*, et al.* (2008). "A prostatic intraepithelial neoplasia-dependent p27 Kip1 checkpoint induces senescence and inhibits cell proliferation and cancer progression." Cancer Cell **14**(2): 146-155.

Makinen, N., Mehine, M., Tolvanen, J.*, et al.* (2011). "MED12, the mediator complex subunit 12 gene, is mutated at high frequency in uterine leiomyomas." Science **334**(6053): 252-255.

McKenna, A., Hanna, M., Banks, E.*, et al.* (2010). "The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data." Genome Res **20**(9): 1297-1303.

Meng, M. V., Elkin, E. P., Latini, D. M.*, et al.* (2005). "Treatment of patients with high risk localized prostate cancer: results from cancer of the prostate strategic urological research endeavor (CaPSURE)." J Urol **173**(5): 1557-1561.

Menon, R., Deng, M., Ruenauver, K.*, et al.* (2013). "Somatic copy number alterations by whole exome sequencing implicates YWHAZ and PTK2 in castration-resistant prostate cancer." J Pathol **231**(4): 505-516.

Metzger, E., Imhof, A., Patel, D.*, et al.* (2010). "Phosphorylation of histone H3T6 by PKCbeta(I) controls demethylation at histone H3K4." Nature **464**(7289): 792-796.

Mimeault, M. and Batra, S. K. (2006). "Recent advances on multiple tumorigenic cascades involved in prostatic cancer progression and targeting therapies." Carcinogenesis **27**(1): 1-22.

Minamiya, Y., Nakagawa, T., Saito, H.*, et al.* (2005). "Increased expression of myosin light chain kinase mRNA is related to metastasis in non-small cell lung cancer." Tumour Biol **26**(3): 153-157.

Minner, S., Enodien, M., Sirma, H.*, et al.* (2011). "ERG status is unrelated to PSA recurrence in radically operated prostate cancer in the absence of antihormonal therapy." Clin Cancer Res **17**(18): 5878-5888.

Nakamura, J., Aoyagi, S., Nanchi, I.*, et al.* (2009). "Overexpression of eukaryotic elongation factor eEF2 in gastrointestinal cancers and its involvement in G2/M progression in the cell cycle." Int J Oncol **34**(5): 1181-1189.

Natarajan, T. G., Kallakury, B. V., Sheehan, C. E.*, et al.* (2010). "Epigenetic regulator MLL2 shows altered expression in cancer cell lines and tumors from human breast and colon." Cancer Cell Int **10**: 13.

Nhili, R., Peixoto, P., Depauw, S.*, et al.* (2013). "Targeting the DNA-binding activity of the human ERG transcription factor using new heterocyclic dithiophene diamidines." Nucleic Acids Res **41**(1): 125-138.

Niggli, V., Schmid, M. and Nievergelt, A. (2006). "Differential roles of Rho-kinase and myosin light chain kinase in regulating shape, adhesion, and migration of HT1080 fibrosarcoma cells." Biochem Biophys Res Commun **343**(2): 602-608.

Norris, J. D., Chang, C. Y., Wittmann, B. M.*, et al.* (2009). "The homeodomain protein HOXB13 regulates the cellular response to androgens." Mol Cell **36**(3): 405-416.

O'Kelly, F., Marignol, L., Meunier, A.*, et al.* (2012). "MicroRNAs as putative mediators of treatment response in prostate cancer." Nat Rev Urol **9**(7): 397-407.

Ostling, P., Leivonen, S. K., Aakula, A.*, et al.* (2011). "Systematic analysis of microRNAs targeting the androgen receptor in prostate cancer cells." Cancer Res **71**(5): 1956-1967.

Oudes, A. J., Roach, J. C., Walashek, L. S.*, et al.* (2005). "Application of Affymetrix array and Massively Parallel Signature Sequencing for identification of genes involved in prostate cancer progression." BMC cancer **5**: 86.

Palanisamy, N., Ateeq, B., Kalyana-Sundaram, S.*, et al.* (2010). "Rearrangements of the RAF kinase pathway in prostate cancer, gastric cancer and melanoma." Nat Med **16**(7): 793-798.

Pan, Y., Kytola, S., Farnebo, F.*, et al.* (1999). "Characterization of chromosomal abnormalities in prostate cancer cell lines by spectral karyotyping." Cytogenet Cell Genet **87**(3-4): 225-232.

Park, K., Tomlins, S. A., Mudaliar, K. M.*, et al.* (2010). "Antibody-based detection of ERG rearrangement-positive prostate cancer." Neoplasia **12**(7): 590-598.

Parkin, D. M., Bray, F., Ferlay, J.*, et al.* (2005). "Global cancer statistics, 2002." CA Cancer J Clin **55**(2): 74-108.

Parmer, T. G., Ward, M. D., Yurkow, E. J.*, et al.* (1999). "Activity and regulation by growth factors of calmodulin-dependent protein kinase III (elongation factor 2-kinase) in human breast cancer." Br J Cancer **79**(1): 59-64.

Paulo, P., Barros-Silva, J. D., Ribeiro, F. R.*, et al.* (2012). "FLI1 is a novel ETS transcription factor involved in gene fusions in prostate cancer." Genes Chromosomes Cancer **51**(3): 240-249.

Peng, X., Guo, W., Liu, T.*, et al.* (2011). "Identification of miRs-143 and -145 that is associated with bone metastasis of prostate cancer and involved in the regulation of EMT." PLoS One **6**(5): e20341.

Perner, S., Demichelis, F., Beroukhim, R.*, et al.* (2006). "TMPRSS2:ERG fusion-associated deletions provide insight into the heterogeneity of prostate cancer." Cancer Res **66**(17): 8337-8341.

Perner, S., Mosquera, J. M., Demichelis, F.*, et al.* (2007). "TMPRSS2-ERG fusion prostate cancer: an early molecular event associated with invasion." Am J Surg Pathol **31**(6): 882-888.

Pflueger, D., Rickman, D. S., Sboner, A.*, et al.* (2009). "N-myc downstream regulated gene 1 (NDRG1) is fused to ERG in prostate cancer." Neoplasia **11**(8): 804-811.

Pflueger, D., Terry, S., Sboner, A.*, et al.* (2011). "Discovery of non-ETS gene fusions in human prostate cancer using next-generation RNA sequencing." Genome Res **21**(1): 56-67.

Prensner, J. R., Iyer, M. K., Balbin, O. A.*, et al.* (2011). "Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression." Nat Biotechnol **29**(8): 742-749.

Quinlan, A. R. and Hall, I. M. (2010). "BEDTools: a flexible suite of utilities for comparing genomic features." Bioinformatics **26**(6): 841-842.

Rajan, P., Dalgliesh, C., Carling, P. J.*, et al.* (2011). "Identification of novel androgen-regulated pathways and mRNA isoforms through genome-wide exon-specific profiling of the LNCaP transcriptome." PLoS One **6**(12): e29088.

Reva, B., Antipin, Y. and Sander, C. (2011). "Predicting the functional impact of protein mutations: application to cancer genomics." Nucleic Acids Res **39**(17): e118.

Reynolds, M. A. (2008). "Molecular alterations in prostate cancer." Cancer Lett **271**(1): 13-24.

Ribas, J., Ni, X., Haffner, M.*, et al.* (2009). "miR-21: an androgen receptor-regulated microRNA that promotes hormone-dependent and hormone-independent prostate cancer growth." Cancer Res **69**(18): 7165-7169.

Rickman, D. S., Pflueger, D., Moss, B.*, et al.* (2009). "SLC45A3-ELK4 is a novel and frequent erythroblast transformation-specific fusion transcript in prostate cancer." Cancer Res **69**(7): 2734-2738.

Rider, J. R., Sandin, F., Andren, O.*, et al.* (2013). "Long-term outcomes among noncuratively treated men according to prostate cancer risk category in a nationwide, population-based study." Eur Urol **63**(1): 88-96.

Robbins, C. M., Tembe, W. A., Baker, A.*, et al.* (2011). "Copy number and targeted mutational analysis reveals novel somatic events in metastatic prostate tumors." Genome Res **21**(1): 47-55.

Robinson, J. T., Thorvaldsdottir, H., Winckler, W.*, et al.* (2011). "Integrative genomics viewer." Nat Biotechnol **29**(1): 24-26.

Rowley, J. D. (2001). "Chromosome translocations: dangerous liaisons revisited." Nat Rev Cancer **1**(3): 245-250.

Sahu, B., Laakso, M., Ovaska, K.*, et al.* (2011). "Dual role of FoxA1 in androgen receptor binding to chromatin, androgen signalling and prostate cancer." EMBO J **30**(19): 3962-3976.

Sampson, N., Neuwirt, H., Puhr, M.*, et al.* (2013). "In vitro model systems to study androgen receptor signaling in prostate cancer." Endocr Relat Cancer **20**(2): R49-64.

Schaefer, G., Mosquera, J. M., Ramoner, R.*, et al.* (2013). "Distinct ERG rearrangement prevalence in prostate cancer: higher frequency in young age and in low PSA prostate cancer." Prostate Cancer Prostatic Dis **16**(2): 132-138.

Schuster-Bockler, B. and Lehner, B. (2012). "Chromatin organization is a major influence on regional mutation rates in human cancer cells." Nature **488**(7412): 504-507.

Selth, L. A., Townley, S., Gillis, J. L.*, et al.* (2012). "Discovery of circulating microRNAs associated with human prostate cancer using a mouse model of disease." Int J Cancer **131**(3): 652-661.

Severi, G., Hayes, V. M., Neufing, P.*, et al.* (2006). "Variants in the prostate-specific antigen (PSA) gene and prostate cancer risk, survival, and circulating PSA." Cancer Epidemiol Biomarkers Prev **15**(6): 1142-1147.

Sharma, N. L., Massie, C. E., Ramos-Montoya, A.*, et al.* (2013). "The androgen receptor induces a distinct transcriptional program in castration-resistant prostate cancer in man." Cancer Cell **23**(1): 35-47.

Shi, X. B., Xue, L., Ma, A. H.*, et al.* (2012). "Tumor suppressive miR-124 targets androgen receptor and inhibits proliferation of prostate cancer cells." Oncogene **32**(35): 4130-4138.

Shi, X. B., Xue, L., Yang, J.*, et al.* (2007). "An androgen-regulated miRNA suppresses Bak1 expression and induces androgen-independent growth of prostate cancer cells." Proc Natl Acad Sci U S A **104**(50): 19983-19988.

Shih, A. H., Abdel-Wahab, O., Patel, J. P.*, et al.* (2012). "The role of mutations in epigenetic regulators in myeloid malignancies." Nat Rev Cancer **12**(9): 599-612.

Siegel, R., DeSantis, C., Virgo, K.*, et al.* (2012). "Cancer treatment and survivorship statistics, 2012." CA Cancer J Clin **62**(4): 220-241.

Spahn, M., Kneitz, S., Scholz, C. J.*, et al.* (2010). "Expression of microRNA-221 is progressively reduced in aggressive prostate cancer and metastasis and predicts clinical recurrence." Int J Cancer **127**(2): 394-403.

Spans, L., Atak, Z. K., Van Nieuwerburgh, F.*, et al.* (2012). "Variations in the exome of the LNCaP prostate cancer cell line." Prostate **72**(12): 1317-1327.

Spans, L., Clinckemalie, L., Helsen, C.*, et al.* (2013). "The genomic landscape of prostate cancer." Int J Mol Sci **14**(6): 10822-10851.

Sramkoski, R. M., Pretlow, T. G., 2nd, Giaconia, J. M.*, et al.* (1999). "A new human prostate carcinoma cell line, 22Rv1." In Vitro Cell Dev Biol Anim **35**(7): 403-409.

Srikantan, V., Zou, Z., Petrovics, G.*, et al.* (2000). "PCGEM1, a prostate-specific gene, is overexpressed in prostate cancer." Proc Natl Acad Sci U S A **97**(22): 12216-12221.

Stephens, P. J., Greenman, C. D., Fu, B.*, et al.* (2011). "Massive genomic rearrangement acquired in a single catastrophic event during cancer development." Cell **144**(1): 27-40.

Stone, K. R., Mickey, D. D., Wunderli, H.*, et al.* (1978). "Isolation of a human prostate carcinoma cell line (DU 145)." Int J Cancer **21**(3): 274-281.

Sun, M., Song, C. X., Huang, H.*, et al.* (2013). "HMGA2/TET1/HOXA9 signaling pathway regulates breast cancer growth and metastasis." Proc Natl Acad Sci U S A **110**(24): 9920-9925.

Sun, R., Fu, X., Li, Y.*, et al.* (2009). "Global gene expression analysis reveals reduced abundance of putative microRNA targets in human prostate tumours." BMC Genomics **10**: 93.

Svensson, M. A., Perner, S., Ohlson, A. L.*, et al.* (2014). "A Comparative Study of ERG Status Assessment on DNA, mRNA, and Protein Levels Using Unique Samples from a Swedish Biopsy Cohort." Appl Immunohistochem Mol Morphol **22**(2): 136-141.

Taatjes, D. J. (2010). "The human Mediator complex: a versatile, genome-wide regulator of transcription." Trends Biochem Sci **35**(6): 315-322.

Tahiliani, M., Koh, K. P., Shen, Y.*, et al.* (2009). "Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1." Science **324**(5929): 930-935.

Taylor, B. S., Schultz, N., Hieronymus, H.*, et al.* (2010). "Integrative genomic profiling of human prostate cancer." Cancer Cell **18**(1): 11-22.

Thalmann, G. N., Anezinis, P. E., Chang, S. M.*, et al.* (1994). "Androgen-independent cancer progression and bone metastasis in the LNCaP model of human prostate cancer." Cancer Res **54**(10): 2577-2581.

Tohtong, R., Phattarasakul, K., Jiraviriyakul, A.*, et al.* (2003). "Dependence of metastatic cancer cell invasion on MLCK-catalyzed phosphorylation of myosin regulatory light chain." Prostate Cancer Prostatic Dis **6**(3): 212-216.

Tomlins, S. A., Laxman, B., Dhanasekaran, S. M.*, et al.* (2007). "Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer." Nature **448**(7153): 595-599.

Tomlins, S. A., Laxman, B., Varambally, S.*, et al.* (2008). "Role of the TMPRSS2-ERG gene fusion in prostate cancer." Neoplasia **10**(2): 177-188.

Tomlins, S. A., Mehra, R., Rhodes, D. R.*, et al.* (2006). "TMPRSS2:ETV4 gene fusions define a third molecular subtype of prostate cancer." Cancer Res **66**(7): 3396-3400.

Tomlins, S. A., Rhodes, D. R., Perner, S.*, et al.* (2005). "Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer." Science **310**(5748): 644-648.

Tranchevent, L. C., Barriot, R., Yu, S.*, et al.* (2008). "ENDEAVOUR update: a web resource for gene prioritization in multiple species." Nucleic Acids Res **36**(Web Server issue): W377-384.

Trapnell, C., Roberts, A., Goff, L.*, et al.* (2012). "Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks." Nat Protoc **7**(3): 562-578.

Trojan, L., Schaaf, A., Steidler, A.*, et al.* (2005). "Identification of metastasis-associated genes in prostate cancer by genetic profiling of human prostate cancer cell lines." Anticancer Res **25**(1A): 183-191.

Vaarala, M. H., Porvari, K., Kyllonen, A.*, et al.* (2001a). "The TMPRSS2 gene encoding transmembrane serine protease is overexpressed in a majority of prostate cancer patients: detection of mutated TMPRSS2 form in a case of aggressive disease." Int J Cancer **94**(5): 705-710.

Vaarala, M. H., Porvari, K. S., Kellokumpu, S.*, et al.* (2001b). "Expression of transmembrane serine protease TMPRSS2 in mouse and human tissues." J Pathol **193**(1): 134-140.

van Bokhoven, A., Varella-Garcia, M., Korch, C.*, et al.* (2003). "Molecular characterization of human prostate carcinoma cell lines." Prostate **57**(3): 205-225.

Van Loo, P., Nordgard, S. H., Lingjaerde, O. C.*, et al.* (2010). "Allele-specific copy number analysis of tumors." Proc Natl Acad Sci U S A **107**(39): 16910-16915.

Veldscholte, J., Ris-Stalpers, C., Kuiper, G. G.*, et al.* (1990). "A mutation in the ligand binding domain of the androgen receptor of human LNCaP cells affects steroid binding characteristics and response to anti-androgens." Biochem Biophys Res Commun **173**(2): 534-540.

Vlietstra, R. J., van Alewijk, D. C., Hermans, K. G.*, et al.* (1998). "Frequent inactivation of PTEN in prostate cancer cell lines and xenografts." Cancer Res **58**(13): 2720-2723.

Volinia, S., Calin, G. A., Liu, C. G.*, et al.* (2006). "A microRNA expression signature of human solid tumors defines cancer gene targets." Proc Natl Acad Sci U S A **103**(7): 2257-2261.

Waltering, K. K., Porkka, K. P., Jalava, S. E.*, et al.* (2011). "Androgen regulation of micro-RNAs in prostate cancer." Prostate **71**(6): 604-614.

Wang, R., Chinnaiyan, A. M., Dunn, R. L.*, et al.* (2009). "Rational approach to implementation of prostate cancer antigen 3 into clinical care." Cancer **115**(17): 3879-3886.

Weischenfeldt, J., Simon, R., Feuerbach, L.*, et al.* (2013). "Integrative genomic analyses reveal an androgen-driven somatic alteration landscape in early-onset prostate cancer." Cancer Cell **23**(2): 159-170.

Wong, O. G., Nitkunan, T., Oinuma, I.*, et al.* (2007). "Plexin-B1 mutations in prostate cancer." Proc Natl Acad Sci U S A **104**(48): 19040-19045.

Wu, C., Wyatt, A. W., Lapuk, A. V.*, et al.* (2012). "Integrated genome and transcriptome sequencing identifies a novel form of hybrid and aggressive prostate cancer." J Pathol **227**(1): 53-61.

Xie, B. X., Zhang, H., Wang, J.*, et al.* (2011). "Analysis of differentially expressed genes in LNCaP prostate cancer progression model." J Androl **32**(2): 170-182.

Xu, W., Yang, H., Liu, Y.*, et al.* (2011). "Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases." Cancer Cell **19**(1): 17-30.

Yang, H., Liu, Y., Bai, F.*, et al.* (2013). "Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation." Oncogene **32**(5): 663-669.

Yegnasubramanian, S., Haffner, M. C., Zhang, Y.*, et al.* (2008). "DNA hypomethylation arises later in prostate cancer progression than CpG island hypermethylation and contributes to metastatic tumor heterogeneity." Cancer Res **68**(21): 8954-8967.

Zhang, X., Cowper-Sal lari, R., Bailey, S. D.*, et al.* (2012). "Integrative functional genomics identifies an enhancer looping to the SOX9 gene disrupted by the 17q24.3 prostate cancer risk locus." Genome Res **22**(8): 1437-1446.

Zhang, Y., Liu, T., Meyer, C. A.*, et al.* (2008). "Model-based analysis of ChIP-Seq (MACS)." Genome Biol **9**(9): R137.